Essays on the Economics of Education

Sandoy, Tróndur Møller

Document Version
Final published version

DOI:
10.22439/phd.18.2023

Publication date:
2023

COPENHAGEN BUSINESS SCHOOL
SOLBJERG PLADS 3
DK-2000 FREDERIKSBERG
DANMARK

WWW.CBS.DK

ISSN 0906-6934

Print ISBN: 978-87-7568-181-5
Online ISBN: 978-87-7568-182-2

DOI: https://doi.org/10.22439/phd.18.2023

ESSAYS ON THE ECONOMICS OF EDUCATION

PhD Series 18.2023

Tróndur Møller Sandoy

ESSAYS ON THE ECONOMICS OF EDUCATION

Department of Economics

PhD Series 18.2023

CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Doctor of Philosophy
Doctoral Thesis in Economics

Copenhagen Business School
Department of Economics

Essays on the Economics of Education

Tróndur Møller Sandoy

Main supervisor: Fane N. Groes
Secondary supervisor: Bertel Schjerning

Tróndur Møller Sandoy
Essays on the Economics of Education

First edition 2023
Ph.D. Series 18.2023

ISSN 0906-6934

Print ISBN: 978-87-7568-181-5
Online ISBN: 978-87-7568-182-2

DOI: https://doi.org/10.22439/phd.18.2023

Acknowledgements

This thesis is the result of my doctoral studies as a PhD fellow at the Department of Economics at Copenhagen Business School. I am grateful for the Department's support during my studies.

First, I want to thank my primary supervisor Fane N. Groes. Thank you for giving me time and encouragement to develop my own research agenda. It has been a pleasure to be supervised by you, and it has been great to have your support and encouragement every step of the way. It has also been a pleasure to coauthor the second Chapter of my thesis with you. I hope to continue working with and learning from you in the coming years. For my secondary supervisor Bertel Schjerning, it was your teaching in the course Dynamic Programming - Theory, Computation, and Empirical Applications, which I took during my master's degree at the University of Copenhagen, which convinced me I wanted to do a PhD. Thank you for your great support during my studies and great comments on my work.

I would also like to thank my other coauthors, Jeanette Walldorf, Edith Madsen, Alexander Fischer, and Andrei Gorshkov. I have learned a lot from you. Jeanette, we started in the same PhD cohort at the Department with similar research interests. It was always great to have someone to talk to at the same stage in their studies. Thank you for being a great colleague and friend. Thank you, Edith, it has been a pleasure to coauthor with you, and I hope to continue our work together. Thank you Alex and Andrei, it has truly been a pleasure to work with you.

I want to thank my current colleagues at the University of the Faroe Islands for offering me a position here before I finished my PhD thesis and for giving me time to finish it. Your support during this time has meant a lot to me.

I also want to thank my former colleagues at the Department of Economics at Copenhagen Business School. Thank you for all your valuable comments during the presentations of my work. A special thanks to Herdis Steingrimsottir and Hans Henrik Sievertsen for their great comments during my predefence.

During my studies, I was lucky enough to be invited to visit the Department of Economics at University College London. I want to thank Professor Aureo de Paula for inviting me and supervising me while I was there.

I want to thank the Novo Nordisk Foundation (grant number NNF16OC0021056) and the Rockwool Foundation for the financial support of my studies.

Chapters 3 and 4 use data on Business Economics students at Copenhagen Business School. Understanding these data and the historical institutional details at Copenhagen Business School took a lot of work. I want to thank Marie Skibsted, Annette J. Hansen, Louise W. Jensen, Anne-Sophie Kvisgaard, Tine B. Poulsen, and Pernille Brandt for help with this.

To my family and friends, thank you for your encouragement and continued support through my studies.

The most special thanks go to my partner Vár. These last years we have experienced some significant changes in our lives. We have been blessed with our two sons, Jónas and Ári, we moved back to the Faroe Islands after having lived in Denmark for many years and bought our first house together. Your support during my studies has meant the world to me, and I do not think I could have done it without you. I look forward to celebrating this accomplishment with you.

English Abstract

This thesis consists of four self-contained chapters on the economics of education. Chapter 1 focuses on the demand for higher education. Chapter 2 studies the effect of absence in Vocational Education and Training. Finally Chapter 3 and 4 are concerned with social interactions. Chapter 3 studies the peer effects in relation to the gender gap for top earners and Chapter 4 studies the effect of alumni networks on labor market outcomes.

In Chapter 1, "Demand for Higher Education and Beliefs on Admission Chances", I study the effects of reducing capacities for higher education programs on the demand by applicants. I model the applicants problem with a portfolio choice model, where I also explicitly model applicants individual beliefs on assignment probabilities. Using detailed application data covering all the applications for higher education in Denmark in 2014, combined with administrative register data I estimate the preferences of applicants for university programs. With the estimated preference parameters I can run policy experiments, where I reduce capacities for some programs. I compare the results from the policy experiments for two different scenarios, a) where applicants can take the changes to capacities into account and b) the current setting where capacities are only partly revealed ex post, and applicants cannot take them into account. I find that applicants change their preferences to take changes in capacities into account when I allow them to update their beliefs. Further, I find that the changes are not one to one with reductions in capacities and that applicants to other programs also changes their applications in response to ripple effects caused by the implemented matching mechanism.

In Chapter 2, "Absence and Completion among students in Vocational Education", which is joint work with Fane N. Groes and Edith Madsen, we analyze the effect of school absence on program completion among students in Vocational Education in Denmark. We use data collected on daily student attendance in combination with register data on graduation and background characteristics. As the student absence is likely endogenous in relation to completion, we use an instrumental variable approach to estimate the causal effect. We use different meteorological weather observations as instruments for students' absence and find that absence has a large and significant negative effect on the probability for completion. We support our results with a second instrument, which is based on the fact that have a panel dimension in student absences, while completion is crossectional.

In Chapter 3, "University Peers and Labour Market Gender Gaps", which is joint work with Alexander Fischer, Andrei Gorshkov, and Jeanette Walldorf, we investigate how university peers affect the divergence in career paths of male and female students at the top of the income distribution. We use university records covering 21 cohorts of in-

coming students that are randomly assigned to tutorial groups in Copenhagen Business School and merge it with high-quality administrative register data on students' careers. We show that females who are assigned to peers of higher ability suffer severe earnings losses, have weaker labor market attachment, are less likely to work in positions with management responsibilities, and to reach the top of the earnings distribution. The negative effect on female careers is more pronounced for high-ability female students and driven by exposure to male peers. To identify mechanisms underlying the effect, we investigate how peer ability shapes education outcomes and fertility behavior. We conclude that the differential reaction of male and female students to academic environments characterized by high-ability peers contributes to labor market gender gaps.

In Chapter 4, "Peers and Careers: Labour Market Effects of Alumni Networks" which is joint work with the same set of coauthors, we study the importance of social connections formed among university peers in terms of shaping their future careers. We use the same data as in Chapter 3. We find that students randomly assigned to the same tutorial group tend to have more similar careers than students from the same cohort but a different tutorial group: they tend to work in the same occupations and industries and are more likely to be hired by the same employer. The strongest "excess" similarities of tutorial group peers over cohort peers are observed at the most disaggregated level, the workplace. This effect is strong, persistent (although decreasing over time), characterized by homophily, and pronounced the most for students from the wealthiest family backgrounds. By comparing the transitions of students to workplaces with incumbent group peers to workplaces with incumbent cohort peers, we find that students benefit from their alumni network by gaining access to more stable and higher-paying jobs.

Keywords: school choice, vocational education and training, social networks, peer effects, education, labour market outcomes

Danish Abstract

Denne afhandling består af fire selvstændige kapitler om uddannelsesøkonomi. Kapitel 1 fokuserer på efterspørgslen efter højere uddannelse. Kapitel 2 undersøger effekten af fravær i løbet af erhvervsuddannelser. Kapitel 3 og 4 omhandler begge sociale interaktioner. Kapitel 3 undersøger effekten af klassekammerater på kønsforskelle for toplønnede, mens Kapitel 4 undersøger effekterne af alumnenetværk på arbejdsmarkedsresultater.

I Kapitel 1, "Demand for Higher Education and Beliefs on Admission Chances", undersøger jeg effekten af at reducere antallet af studiepladser på universitetsuddannelser på efterspørgslen efter universitetsuddannelserne. Jeg modellerer ansøgerens problem med en portefølje model, hvor jeg også eksplicit modellerer ansøgeres individuelle subjektive sandsynligheder for at blive optaget på en given uddannelse. Ved brug af detaljerede data på ansøgninger i 2014 kombineret med administrative register data, estimerer jeg ansøgernes præferencer for universitetsuddannelser. Vedhjælp af de estimerede præferenceparametre og modellen kan jeg udføre politikforsøg, hvor jeg reducerer antallet af studiepladser for nogle uddannelser. Jeg sammenligner mine resultater fra politikforsøgene for to forskellige scenarier, a) hvor ansøgere har mulighed for at tage højde for ændringerne i antal studiepladser og b) den nuværende sammenhæng hvor antal studiepladser kun delvist bliver afsløret efter at ansøgere har fået at vide om de er optaget eller ej, og at de derfor ikke har mulighed for at tage højde for ændringer. Jeg finder at ansøgere ændrer deres ansøgninger for at tage højde for ændringerne i antal studiepladser, når jeg giver dem mulighed for at opdaterer deres subjektive sandsynligheder. Videre finder jeg også at ændringer in deres ansøgninger ikke er en til en med ændringen i antal studiepladser og at ansøger til andre upåvirkede uddannelser også ændrer i deres ansøgninger som reaktion på ringvirkninger forårsaget af den implementerede matchningsmekanisme.

I Kapitel 2, "Absence and Completion among students in Vocational Education", som er skrevet sammen med Fane N. Groes og Edith Madsen, analyserer vi effekten af skolefravær på sandsynligheden for at fuldføre en uddannelse for elever på erhvervsskoler i Danmark. Vi anvender indsamlede data på fravær på dagsbasis sammen med register data med informationer om fuldførelse af uddannelser og baggrundsinformationer. Da elevernes fravær sandsynligvis er endogent i forhold til fuldførelse, bruger vi en instrumentel variabel tilgang til at estimere årsagseffekten. Vi anvender forskellige meteorologiske opmålinger som instrumenter for elevernes fravær og finder at fravær har en stor og signifikant negativ effekt på sandsynligheden for at fuldføre uddannelsen. Vi underbygger vores resultater med en anden instrument variabel tilgang. Den er baseret på at

vi har en paneldimension i elevernes fravær, mens vi har tværsnitsdata på fuldførelse.

I Kapitel 3, ”University Peers and Labour Market Gender Gaps”, som er skrevet sammen med Alexander Fischer, Andrei Gorshkov og Jeanette Walldorf, undersøger vi hvordan klassekammerater påvirker afvigelsen i mandlige og kvindelige studerendes karriereforløb i toppen af indkomstfordelingen. Vi anvender data fra universitetsregistre, der dækker 21 årgange af nystartede studerende på Copenhagen Business School, der er tilfældigt fordelt på klasser. Vi sammensætter dette med administrative registerdata for at få oplysninger om studerendes karriere. Vi viser at kvinder som kommer i klasse med klassekammerater med høje akademiske evner lider alvorlige indtægtstab, har en svagere tilknytning til arbejdsmarkedet, er mindre tilbøjelige til at arbejde i stillinger med ledelsesansvar og at nå toppen af indkomstfordelingen. De negative effekter på kvinders karriere er mest udtalte for kvindelige studerende med høje akademiske evner og det er mandlige klassekammerater, der driver effekterne. For at identificere mekanismer, der ligger til grund for effekten, undersøger vi, hvordan klassekammeraters-evner former uddannelsesresultater og fertilitetsadfærd. Vi konkluderer, at mandlige og kvindelige studerendes forskellige reaktion på akademiske miljøer præget af dygtige jævnaldrende bidrager til kønsforskelle på arbejdsmarkedet.

I Kapitel 4, ”Peers and Careers: Labour Market Effects of Alumni Networks”, som er skrevet sammen med de samme medforfattere, undersøger vi betydningen af sociale forbindelser imellem studiekammerater under højere uddannelse i forhold til elevernes fremtidige karrierer. Vi anvender det samme data som i Kapitel 3. Vi finder at studerende som tilfældigt er placeret i samme klasser, er tilbøjelige til at have karrierer som ligner hinanden mere end studerende fra samme årgang, som tilfældigt er placeret i forskellige klasser: De har en tendens til at arbejde i de samme erhverv og industrier og de har også en højere sandsynlighed for at blive ansat af samme arbejdsgiver. De stærkeste ”overskydende” ligheder mellem klassekammerater i forhold til studerende i samme årgang, men forskellige klasser, observeres på det mest opdelte niveau, arbejdspladsen. Denne effekt er stærk, vedvarende (selvom den aftager over tid), præget af homofili og mest udtalt for studerende fra de rigeste familiemæssige baggrunde. Ved at sammenligne tilgangen til arbejdspladser med allerede etablerede klassekammerater med tilgangen til arbejdspladser med allerede etablerede årgangskammerater, finder vi, at de studerende drager fordel af deres netværk ved at få adgang til mere stabile og bedre betalte stillinger.

Nøgleord: valg af skole, erhvervsskoler, sociale netværk, peer-effekter, uddannelse, arbejdsmarkedsresultater

Contents

Introduction

This thesis contributes to various strands of economic literature within the overarching themes of the economics of education and labour economics. While Chapter 2 concerns vocational education, the other chapters concern higher education. Chapter 1 contributes to the literature on higher education demand, while Chapters 3 and 4 delve into post-higher education labor market effects. Chapters 3 and 4 combine quasi-experimental data with administrative register data to study the effects of social interactions. Chapter 3 examines the differential effect of class peers on males and females for post-higher education labor market results. In contrast, Chapter 4 explores the labor market effects of networks formed during higher education. By utilizing various econometric methods and modelling techniques, this thesis aims to provide novel insights into these complex relationships, building upon and expanding the curent body of work in these areas.

A small but growing body of work has studied the effect of school absence on educational outcomes and finds that absence in general has a negative impact on student's educational outcomes (Aucejo & Romano, 2016; Cattan et al., 2023; Liu et al., 2019). Further, a related literature studies the effect of the number of school days or the amount of instruction time on educational outcomes, and find a positive effect on test scores and educational attainment (see, e.g., Marcotte (2007), Hansen (2011), Lavy (2020) and Fitzpatrick et al. (2011)). What is also in common for these bodies of work is that they concern elementary or high school.

Chapter 2, together with Fane N. Groes and Edith Madsen, contributes to these bodies of work by studying the effect of absence on educational outcomes in a vocational education and training context (VET). The small existing economic literature on VET has mostly concentrated on labor market outcomes (see, e.g., Hanushek et al., 2017, Hampf and Woessmann, 2017, Bertrand et al., 2021, and Silliman and Virtanen, 2022). The VET area is an understudied area of potentially high policy interest. Firstly, the VET educations in Denmark have a low completion rate, and secondly, students who dropout have a high probability of ending up as unskilled (Groes et al., 2021).

In Chapter 2 we add to the existing studies on VET education by looking at the effect of absence on the probability of completion. We find that absence has a large negative effect on the probability of completion. Our findings suggest that a policy aimed at increasing student attendance can have large effects on the low graduation rates.

Centralized admissions systems to help match applicants with schools or higher education programs are implemented worldwide. A large and growing body of work looks into the interplay between the implemented matching mechanism and the final matches as well as the effect they have on the application behaviour, e.g. through incentivizing strategic behaviour when applicants form their applications (Abdulkadiroglu & Sönmez, 2003; Agarwal & Somaini, 2020; Balinski & Sönmez, 1999; Patnaik et al., 2021). Particularly many papers have looked into the latter. A recent body of work has found that in many of the real world implementations of the otherwise strategy-proof Deferred Acceptance mechanism by Gale and Shapley (1962) for the higher education market, applicants do not report their preferences truthfully (Artemov et al., 2020; Fack et al., 2019; Haeringer & Klijn, 2009; Hassidim et al., 2016). Given this, there is much evidence of the importance of matching mechanisms for ensuring that the matchings between applicants and programs are the best possible. However, there is a lack of evidence on another important feature of the matching market for higher education: the effects of supply changes on applicant demand.

In Chapter 1, I use detailed application data for higher education programs in Denmark combined with administrative register data to study how changes to program capacities affect applicant demand. To address the fact that it can be optimal for applicants not to report their preferences truthfully, I model applicants' demand with a Portfolio Choice model based on Chade and Smith (2006) and explicitly allow applicants to take their beliefs about their admission chances into account. Specifically, I simulate the effects of reducing capacities for programs within humanities.

I find that reducing capacities for programs in humanities while increasing capacities in other programs correspondingly causes applicants to change their applications. Applicants who formerly had programs in humanities in their applications shift to include other programs. At the same time, other applicants also change their applications both because of increases in capacities but also because of ripple effects. Interestingly, although applicants shift away from humanities programs to avoid being unmatched, they are also at the bottom of the GPA distribution, so they still end up being unmatched to a high degree. Further, applicants left unmatched before the changes to capacities can take advantage of the fact that many programs have increased capacities and are, therefore, able to find matches after the change.

Another contribution of the thesis is in terms of understanding the gender gap. Although a lot has happened during the last decades, there still exists a substantial gender gap in most societies across the globe (Blau & Kahn, 2017; Goldin, 2014; Olivetti & Petrongolo, 2016). Further, the most pronounced underrepresentation of women is at the top of the earnings distribution (Bertrand, 2018). There is also a recent body of work that shows that social interactions in educational institutions affect educational choices and outcomes of students (Brenøe & Zölitz, 2020; Feld & Zölitz, 2017) and that these effects can be different for males and females (Feld & Zölitz, 2018; Fischer, 2017; Mouganie & Wang, 2020). Despite this, there is not much evidence of the contribution of social interactions in educational institutions to the gender gap. A large reason for this is a lack of data and credible identification; it is very seldom that researchers have access to both high-quality labor market observations and a convincing strategy to identify peer effects.

Chapter 3, together with Alexander Fischer, Andrei Gorshkov, and Jeanette Walldorf, studies the effect of social interactions during higher education on post-graduate labor market outcomes. Specifically, we estimate the effect of having high-achieving peers on labor market outcomes. We use observations on Business Economics students at Copenhagen Business School, who are randomly allocated to classes when they enroll. We are further able to combine our data with high-quality Danish administrative register data which allows us to follow students up to 20+ years after they have graduated. This provides us with a unique setting to study the effect of university peers on the gender gap for top earners, namely both a credible way to identify peer effects and observations on students' labor market outcomes.

We find that peer ability composition has a lasting effect on labor market outcomes. Importantly while we find little evidence that males are adversely affected, female students' careers are severely negatively affected by exposure to peers of higher average ability. Women exposed to high-achieving peers experience a substantial earnings penalty, are less likely to work in jobs with management responsibilities, and to reach the top of the income distribution. The effects are most pronounced in the long term. Further, we find that high-ability males drive the negative career effects for females. We also explore some potential mechanisms and find that affected female students are more likely not to continue in a master's degree and graduate with lower GPAs. Secondly, we also find that affected female students have children at an earlier age. This chapter thereby sheds light on one source of the remaining gender gap, which works through the differential effects of social interactions during higher education.

Another strand of the literature on social interactions studies the importance of these in terms of networking effects. It is widely believed that who you know can be just as important, if not more, than what you know. A recent body of work shows increasing evidence for the importance of networks in terms of careers (Bayer et al., 2008; Dustmann et al., 2016; Kramarz & Skans, 2014). Although there is evidence of the importance of networks in terms of careers, and there is evidence of the effects of social interactions during education (Carrell et al., 2009; Feld & Zölitz, 2017; Lyle, 2007; Sacerdote, 2001; Zimmerman, 2003), it is, however, ambiguous how to interpret them in the context of social interactions during higher education. On the one hand, the interactions could be limited to the educational context and still lead to similar labor market outcomes, e.g., by affecting educational decisions or simply by capturing the fact that different educational programs attract student types with preferences for certain career opportunities. On the other hand, the interactions could extend into the students' professional lives, where they rely on the networks formed during their studies to get information on jobs and to further their careers.

In Chapter 4, also together with Alexander Fischer, Andrei Gorshkov, and Jeanette Walldorf, we try to bridge this gap. To do this, we rely on the same data and setting as in Chapter 3. We use a dyadic regression framework to measure the excess similarities in career outcomes for students in the same class compared to students in the same cohort but in different classes. We find the strongest effects for students working in the same workplace, which is the most disaggregated level. We do several robustness checks of our results and find that they strongly suggest that the effect we are measuring supports the interpretation that we are measuring contemporaneous alumni interactions, which students use to further their careers.

In a second part of the chapter, we provide evidence that social interactions benefit students' careers. We find that students who join their class peers experience high wage increases and gain access to more stable and higher-paying jobs than students who join their cohort peers.

References

Abdulkadiroglu, A., & Sönmez, T. (2003). School Choice: A Mechanism Design Approach. *American Economic Review*, *93*(3), 729–747.

Agarwal, N., & Somaini, P. (2020). Empirical Analysis of School Assignment Models. *Annual Review of Economics*, *12*(1), 471–501.

Artemov, G., Che, Y.-K., & He, Y. (2020). Strategic 'Mistakes': Implications for Market Design Research. *Working Paper.*

Aucejo, E. M., & Romano, T. F. (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, *55*, 70–87. https://doi.org/10.1016/j.econedurev.2016.08.007

Balinski, M., & Sönmez, T. (1999). A Tale of Two Mechanisms: Student Placement. *Journal of Economic Theory*, *84*(1), 73–94. https://doi.org/10.1006/jeth.1998.2469

Bayer, P., Ross, S., & Topa, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy*, *116*(6), 1150–1196.

Bertrand, M. (2018). Coase Lecture – The Glass Ceiling. *Economica*, *85*(338), 205–231.

Bertrand, M., Mogstad, M., & Mountjoy, J. (2021). Improving educational pathways to social mobility: Evidence from norway's reform 94. *Journal of Labor Economics*, *39*(4), 965–1010.

Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, *55*(3), 789–865.

Brenøe, A., & Zölitz, U. (2020). Exposure to More Female Peers Widens the Gender Gap in STEM Participation. *Journal of Labor Economics, 38*(4).

Carrell, S., Fullerton, R. L., & West, J. (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, *27*(3), 439–464.

Cattan, S., Kamhöfer, D. A., Karlsson, M., & Nilsson, T. (2023). The Long-term Effects of Student Absence : Evidence from Sweden. *The Economic Journal*, *133*, 888–903.

Chade, H., & Smith, L. (2006). Simultaneous Search. *Econometrica*, *74*(5), 1293–1307.

Dustmann, C., Glitz, A., Schönberg, U., & Brücker, H. (2016). Referral-based Job Search Networks. *Review of Economic Studies*, *83*(2), 514–546.

Fack, G., Grenet, J., & He, Y. (2019). Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions. *American Economic Review*, *109*(4), 1486–1529.

Feld, J., & Zölitz, U. (2017). Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. *Journal of Labor Economics*, *35*(2), 387–428.

Feld, J., & Zölitz, U. (2018). *Peers from venus and mars – higher-achieving men foster gender gaps in major choice and labor market outcomes* (Working Paper). In: Cesifo Area Conferences : Economics of Education.

Fischer, S. (2017). The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence. *Labour Economics*, *46*, 211–226.

Fitzpatrick, M. D., Grissmer, D., & Hastedt, S. (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review*, *30*(2), 269–279. https://doi.org/10.1016/j.econedurev.2010.09.004

Gale, D., & Shapley, L. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, *69*(1), 9–15.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091–1119.

Groes, F. N., Madsen, E., & Sandoy, T. M. (2021). *Completion from vocational educations: A register based analysis* (tech. rep.). Copenhagen Business School.

Haeringer, G., & Klijn, F. (2009). Constrained school choice [Publisher: Elsevier Inc.]. *Journal of Economic Theory*, *144*(5), 1921–1947. https://doi.org/10.1016/j.jet.2009.05.002

Hampf, F., & Woessmann, L. (2017). Vocational vs. general education and employment over the life cycle: New evidence from PIAAC. *CESifo Economic Studies*, *63*(3), 255–269. https://doi.org/10.1093/cesifo/ifx012

Hansen, B. (2011). School Year Length and Student Performance: Quasi-Experimental Evidence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2269846

Hanushek, E. A., Schwerdt, G., Woessmann, L., & Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the lifecycle. *Journal of Human Resources*, *52*(1), 48–87. https://doi.org/10.3368/jhr.52.1.0415-7074R

Hassidim, A., Romm, A., & Shorrer, R. I. (2016). "Strategic" Behaviour in a Strategy-Proof Environment. *Working Paper*.

Kramarz, F., & Skans, O. (2014). When Strong Ties are Strong: Networks and Youth Labour Market Entry. *Review of Economic Studies*, *81*(3), 1164–1200.

Lavy, V. (2020). Expanding School Resources and Increasing Time on Task: Effects on Students' Academic and Noncognitive Outcomes. *Journal of the European Economic Association*, *18*(1), 232–265. https://doi.org/10.1093/jeea/jvy054

Liu, J., Lee, M., & Gershenson, S. (2019). The Short-and Long-Run Impacts of Secondary School Absences. *IZA Discussion Papers*, *No. 12613*.

Lyle, D. S. (2007). Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point. *The Review of Economics and Statistics*, *89*(2), 289–299.

Marcotte, D. E. (2007). Schooling and test scores: A mother-natural experiment. *Economics of Education Review*, *26*(5), 629–640. https://doi.org/10.1016/j.econedurev.2006.08.001

Mouganie, P., & Wang, Y. (2020). High Performing Peers and Female STEM Choices in School. *Journal of Labor Economics*, *38*(3), 805–841.
Olivetti, C., & Petrongolo, B. (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics*, *8*, 405–434.
Patnaik, A., Wiswall, M., & Zafar, B. (2021). The routledge handbook of the economics of education, ch. College Majors. Routledge. https://doi.org/10.4324/9780429202520-16
Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, *116*(2), 681–704.
Silliman, M., & Virtanen, H. (2022). Labor market returns to vocational secondary education. *American Economic Journal: Applied Economics*, *14*(1), 197–224.
Zimmerman, D. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *The Review of Economics and Statistics*, *85*(1), 9–23.

CHAPTER 1

Demand for Higher Education and Beliefs on Admission Chances

Abstract

This paper studies the effect of changes to higher education supply on applicant demand in the Danish higher education market, which relies on a Deferred Acceptance type mechanism to match applicants and programs. To do this, I model the applicant's problem with a portfolio choice model and explicitly model applicants' subjective beliefs on admission chances. I estimate the model's parameters using detailed application data from 2014 combined with high-quality register data, allowing me to generate rich program characteristics. Having estimated the parameters of the model, I can then run counterfactual policy experiments where I reduce capacities for some programs to see how applicants respond. I find that, in general, applicant demand responds to changes in supply and that applicants to other programs than the directly affected also change their application behavior.

1.1 Introduction

Worldwide, Centralized admission systems are widely used to match students with schools and higher education programs[1]. Studies show that the allocation mechanism used by the Centralized Admission Systems (here on CAS) affects the welfare and other aspects of the allocations (Abdulkadiroglu & Sönmez, 2003; Balinski & Sönmez, 1999).

Much of this literature looks at how applicant demand responds to changes to the admission systems. In particular, many studies compare the applicant welfare under a manipulable allocation mechanism where applicants have a clear incentive to be strategic in their applications to a counterfactual strategy-proof mechanism, e.g., a Deferred Acceptance (here on DA) type algorithm similar to the one proposed by Gale and Shapley (1962). There are, however, few studies on how student demand responds to changes in

[1] I refer to major and university combinations as programs, e.g., economics at the University of Copenhagen.

supply. The answer is trivial when the mechanism is strategy-proof and applicants report their preferences truthfully. In that case, any changes in supply will only affect demand through changes in the characteristics of applicants and programs. Recent studies show that in many of the real world implementations of the deferred acceptance mechanism, with for example caps on the length of applications that can be submitted, the strategy-proofness result breaks down even though the mechanism itself is non-manipulative (Artemov et al., 2020; Fack et al., 2019; Haeringer & Klijn, 2009; Hassidim et al., 2016). It is not a priori clear how demand will respond to changes to supply when applicants do not report their preferences truthfully.

In this paper, I study how the demand for higher education responds to changes in supply, through changes in program capacities, in an admission system that uses a DA type algorithm where strategy-proofness fails. Further, I estimate applicant preferences for higher education programs. I base the estimation on Danish higher education application data combined with detailed administrative register data. The Danish CAS uses a DA type matching mechanism, similar to the one proposed by Gale and Shapley (1962) with some modifications. Most importantly, applicants cannot include more than eight programs in their applications and because of this constraint it is not necessarily optimal for applicants to report their preferences truthfully (Artemov et al., 2020; Fack et al., 2019; Haeringer & Klijn, 2009; Hassidim et al., 2016). In practice it means that applicants can choose to *skip the impossible* or *leave out not good enough* programs, such that the observed applications are a subset of the applicants true preferences. I report suggestive evidence showing that applicants to a large extent include programs which have GPA cutoffs close to their high school GPA and that they are reporting fewer programs than the limit of eight programs. If applicants reported their preferences truthfully I would expect them to include more programs and also to include programs with higher cutoffs as this should not affect their likelihood of receiving an offer from the other reported programs. The canonical model for applicant preferences assumes that applicants report their preferences truthfully, and estimates of preferences are therefore biased if this assumption fails. To avoid the mentioned bias and to be able to simulate counterfactual policy experiments where I change the supply through program capacities, I need to take the strategizing behavior of applicants into account.

To do this, I set up a portfolio choice model based on the framework in Chade and Smith (2006). The model allows me to relax the assumption of truth-telling by considering applicants' beliefs about assignment probabilities. This allows me to rationalize the observed application behaviour. I apply the bootstrap estimator suggested by Agarwal and Somaini (2018) combined with an assumption on rational beliefs on admission chances to estimate beliefs. I then combine these with detailed individual application data containing all applications in Denmark for 2014 and high-quality administrative register data to estimate applicant beliefs and preferences.

I further include a belief updating channel in my model. This in combination with assuming that preference parameters are policy invariant allows me to perform policy experiments to evaluate how applicant demand responds to changes in program capacities through applicants beliefs. In practice I use the portfolio choice model to solve for the optimal applications given a set of estimated parameters and re-estimating beliefs under a set of counterfactual capacities. I repeat these steps until beliefs have converged, and by comparing the simulated applications using the baseline beliefs with the simulated applications using the updated beliefs I can see how the model predicts demand will respond to a change in capacities. I compare the search patterns of students under the different policy experiments, and as I know the assignment mechanism, the final allocations of students for different capacities. In further also compare the final allocations in the policy experiments with a scenario where applicants do not have information on the changes to the capacities, and therefore are not able to update their beliefs. The changes in capacities that I consider are ”neutral”, as I redistribute any reductions in capacities for affected programs to unaffected programs according to the size of their prior capacities. The reason is that I only model the intensive margin of the application behavior (e.g., the decision for which program(s) to apply). As I do not model the extensive margin (e.g., whether to apply or not), the model is not well suited to measure the impact of an overall change in capacities.

I find that reducing capacities for programs in humanities while increasing capacities for other programs causes applicants to change their application behavior, if they can update their beliefs. I see changes in application behavior even if I condition on whether applicants applied to programs before the change, who either have decreased or increased capacities now. However, applicants with an application to programs in humanities, before the change, shift away from humanities to a higher degree than other applicants shift away from their original field. Further, the characteristics of the programs that applicants with an application to programs in humanities, before the change, shift to are more similar to the programs of other applicants on average. Importantly I also find that revealing capacities and changes to capacities to applicants before they submit their applications might help applicants who otherwise end up being rejected.

This paper’s contributions are multiple and relate to several different strands of the literature. Firstly, I contribute to the large and growing body of work on estimated preferences for higher education. Internationally many papers have estimated preferences for higher education, see for example Patnaik et al. (2021) for a recent survey of the literature. Particularly the contribution of this paper is estimating preferences based on rich measures of program characteristics generated from register data while relaxing the strong assumption of truthfulness by allowing applicants to be strategic in my model. Another way of relaxing the assumption of truthfulness is to use the estimator proposed by Fack et al. (2019). I do not use this approach as I need to fully model the application behaviour to be able perform the policy experiments. The body of work on estimating preferences for higher education is closely related to the literature studying school choice. The two problems are similar in many aspects although they differ in some important ways. Firstly schools mainly differ in the quality of the teaching and the distance to the students home, while university programs include a third dimension in the content which is taught (e.g. business or medicine). For university programs the third dimension is most likely the most important. Further, for university programs it is the student who makes the decision for which programs to apply for, while it is the parents of the student who make the decision in school choice. The seminal paper by Abdulkadiroglu and Sönmez (2003) framed the school choice problem as a mechanism design problem. They analyzed some of the existing allocation mechanisms and offer two alternative mechanisms which provide solutions to some existing problems. The paper inspired a large body of work to better understand the school choice problem and the effects of the allocation mechanisms. Agarwal and Somaini (2020) survey the recent methods to estimate preferences for schools and gives an overview of the empirical results.

Another branch of the literature on school choice is concerned with estimating preferences for schools or higher education programs when applicants do not report their preferences truthfully. Recently several papers have looked into this problem by looking at subjective beliefs on admission chances and information. Agarwal and Somaini (2020) study a model of school choice where applicants have beliefs of admission chances. In their main specification they assume that applicants have rational expectations when forming their beliefs. This approach assumes that applicants have full information when forming their beliefs. I use the same approach to estimate beliefs in my paper. Kapor et al. (2020) elicit beliefs through a survey and find that the beliefs are not in line with an assumption of rational expectations. They relax the assumption of rational expectations by allowing applicants to make mistakes when forming their beliefs. My current model does not allow me to consider mistakes in the formation of beliefs.[2] Chen and He (2021, 2022) show that information costs can affect which preferences applicants report as they might not have enough information on all schools/programs and acquiring information can be costly. I do not properly model information costs in my model, but include a fixed information cost from including additional programs in the application. The reason is that it is difficult to separate the effect of beliefs and information costs on reported preferences.

Secondly, I contribute to the body of work which structurally models applicants' behavior in the school choice problem. Some other papers who model the applicants decision based on the portfolio choice model by Chade and Smith (2006) are Larroucau and Rios (2020), Larroucau and Rios (2022), Ekbatani (2022), and Ajayi and Sidibe (2015). Larroucau and Rios (2020), Larroucau and Rios (2022), and Ekbatani (2022) also model applications to higher education programs in Chile and Iran, while Ajayi and Sidibe (2015) set up at portfolio choice model to study school choice in Ghana. Calsamiglia et al. (2017) set up a structural model which they solve by backward induction for school choice in Barcelona.

Thirdly, I contribute to the small body of work on the effects of changes to the supply of higher education programs, through changes to capacities. My contribution is developing a model framework that allows the researcher to understand how the demand for higher education programs responds to intensive margin changes to the supply through changing capacities in a centralized application system using a deferred acceptance type or similar matching mechanisms. Another paper which studies this is Gandil (2022), who estimates the effect of supply changes on potential earnings, while holding preferences fixed. Gandil (2022) finds that changing supply with 10 slots leads

[2] If applicants make mistakes when forming their beliefs it can bias my preference estimates, although it is difficult to say without having elicited beliefs.

to 15 applicants moving and that it explains 40 percent of the variation in earnings. This body of work is also related to the recent papers on field of study and earnings. Kirkeboen et al. (2016) exploit the implied randomness which arises close the the GPA cutoffs in university allocation mechanism which use a DA type mechanism, to identify the causal effect of crossing the threshold on earnings. Daly et al. (2022) build on this by looking at the effect of having first and second ranked programs in the same broad field versus having these in different broad fields. They find that only students with the first and second ranked programs in different broad fields, are negatively affected by not being accepted in their first ranked program.

The closest papers to this paper are Gandil (2022) and Larroucau and Rios (2020). While Gandil (2022) studies substitution effects that arise due to changes in program capacities for the Danish higher education market, as I do, my paper differs in some important ways. Firstly, he assumes that applicants report their preferences truthfully and that they are not affected by supply, such as e.g. Abdulkadiroglu and Sönmez (2003) and Azevedo and Leshno (2016). I relax the assumption on applicants reporting their preferences truthfully by explicitly modelling the applicants' behavior. His approach allows him to look at the substitution effects from changes to capacities for the whole application market, while I am currently limited to looking at a subsample of the pool of applicants as I do not estimate the preferences for the full population of applicants. Secondly, his focus is on showing the effect on foregone expected earnings caused by changes to supply, while my focus is on understanding how applicant demand responds to changes in supply. My scenario, where I do not allow applicants to update their beliefs, is equivalent to his approach. If I had estimated preferences for the full population, my approach could therefore be seen as containing his approach as a special case. Larroucau and Rios (2020) also develop a Portfolio Choice model to estimate applicant preferences for higher education programs. They do this for the Chilean higher education market and compare the estimates from their model to estimates obtained from a model which assumes strict truth-telling and find that these are biased. I use the same approach as they do to model the applicants' preferences. However, my approach differs as I adapt the model by adding a framework that allows applicants to update their beliefs in response to changes in supply. The mechanism I have implemented for updating applicant beliefs is similar to the mechanism in Larroucau and Rios (2022) but differs in its focus. They use it to update applicants' beliefs about their abilities in a dynamic setting, whereas I use it to update applicants' beliefs after changing capacities.

The remainder of the paper is organized as follows. Section 2.2 describes the Institutional settings and, in particular, the application system in Denmark. Section 2.3 describes the different data sources I use. Section 1.4 describes the application behavior and the characteristics of the programs in the observed applications. In section 1.5, I set up the Portfolio Choice model. Section 1.6 lays out the identification strategy. In section 1.7, I describe the approaches used to estimate beliefs and preference parameters. Section 1.8 presents and discusses the results and policy experiments. Finally, section 1.10 concludes with policy recommendations and suggestions for future research.

1.2 Institutional settings

In this section, I describe Denmark's higher education system, the centralized admission system, and the mechanism used to assign applicants to higher education programs.

1.2.1 Higher Education System

The higher education system in Denmark covers all educations post high-school degrees, e.g., education degrees offered by universities and business academies.

All higher education in Denmark is free, and students receive a stipend while enrolled.[3] Further, students in higher education in Denmark are also eligible for cheap student loans, where they are allowed to uptake a new loan corresponding to the yearly stipend each year with a very favorable repayment scheme. Students also receive a higher tax deduction.

Contrary to the higher education system in, e.g., the US, the Danish higher education system requires students to choose a major when applying, and most courses, except for some possible electives, will be within the subject of the chosen major. However, most programs also allow students to take courses offered by other programs, although there are often strict program-specific rules in place for the contents of these courses. Further, most students who graduate with a bachelor's degree also take a master's degree, according to a report by DST (2016) 83% of students who graduated with a bachelors degree in 2016 chose to enroll in a master's degree program in the same year.

1.2.2 Application system and mechanism

The CAS in Denmark handles virtually all applications for higher education. Applicants submit a Rank Ordered List (here on ROL) with their preferred programs to the CAS, which uses a student proposing DA type mechanism to match applicants and programs. By DA type mechanism, I mean that the mechanism is similar to the one proposed by Gale and Shapley (1962), with some important modifications. The first is an upper limit on how many programs each applicant can include in her application, and it is very common in real-world implementations of the DA mechanism. For example, in Denmark, the limit for the length of applications is set at 8 programs. In other countries, it typically does not vary much from this number with some outlier countries, e.g., it is 10 in Norway (Kirkebøen, 2012), 10 in Chile (Larroucau & Rios, 2020), while it is much higher in Iran, 100, (Ekbatani, 2022).

A second modification is splitting the program capacities into multiple quotas. In 2014, for example, the number of quotas was 70. The most used quotas are quota 1,

[3] Students can receive it up to the normalized study time, e.g., five years for economics and six years for medicine, plus one additional year. The stipend comes with some conditions, e.g., how many ECTS points students have to be enrolled in yearly, although these conditions have changed over the years. In 2014 the stipend amount was 5,839 DKK before taxes.

quota 2, and quota 1 and 2 standby.[4] The applicants who apply for a given program through quota 1 are evaluated by their high school GPA and possibly some program-specific requirements, for example, a requirement to have passed a certain level of some set course in high school or a grade above some threshold in a certain course in high school or a program specific minimum GPA threshold. Quota 1 constitutes the majority of all applications as well as capacities (offers). Applicants who apply through quota 2 are evaluated on other measures than their high school GPA, e.g., grades or levels from specific courses in high school, motivational letters, relevant past work experience, and so forth, which differ across programs. It is, however, also important to note that all applicants who apply through quota 2 are first tested using the quota 1 criteria before being tested on the quota 2 criteria, if applicable (the student has a high school degree and the program's admission is not solely through quota 2). Lastly, applicants can also cross off the standby (either quota 1 or quota 2) option in combination with either quota 1 or quota 2. Programs where the applicant has marked standby, are evaluated on the same requirements as pure quota 1 or 2. If an applicant is rejected on either of these "main" quotas for a given program the applicant is evaluated on the standby requirements for the standby capacity for the given program. If the applicant is offered a standby seat in a program, she is offered admission if enough applicants reject their offer. If not, she is guaranteed admission to this program next year. Being accepted in standby, however, also means that the applicant is not evaluated for any programs she listed as a lower priority in her application.

Most programs have quota 2 capacities ranging up to 10% of their total capacities. Further, the number of seats allocated to standby on quota 1 and 2 is also small compared to the total capacities.

Deadlines for applications differ by type of quota. For example, the deadline for submitting through quota 2 is on the 15th of March, and the deadline for submitting through quota 1 is on the 5th of July.

After all offers are given, applicants choose to accept or reject the offer. After that, the aftermarket begins. Applicants who were accepted through the standby quota are offered seats in the given program if rejected offers have freed up any. After that, applicants who were either not given an offer or rejected their offer and new applicants can apply for all programs with free capacities. The applications through the aftermarket go directly to the different programs. I ignore the aftermarket in this paper.

1.3 Data

In this section, I outline the different data sources used, detail how I generate the program-specific characteristics, and state the selection criteria used to define the samples used in the analysis.

[4]The other quotas are, e.g., international students from non-Scandinavian countries and applicants from Greenland, among others. These make up an insignificant number of applicants and capacities, and I ignore them to simplify the analysis.

1.3.1 Application data, capacities, and specific requirements

The primary data source is detailed information on submitted higher education applications in Denmark. I have access to all submitted applications to the Centralized Admissions System in Denmark for 1993-2015. This data contains demographic and educational background information on all applicants (GPA used for application, age, sex, citizenship, high school, type of high school, and other relevant information) along with information on which programs and, importantly, in what specific order each applicant has included these programs in her application and which program she received an offer from if any. The GPA measure in the application data includes GPA multipliers.[5] The application data also contains detailed information for each ranked program on through which quota the application is. The raw dataset contains 3,385,555 applicants × program observations. I restrict the data to applications made in 2014, which leaves me with 244,198 applicants × program observations before I make any selection. There are three distinct reasons for only using applications in 2014. Firstly, I want data that is recent to be able to inform current policy. Secondly, I chose 2014 as I need a lot of data on graduated students to generate some program-specific characteristics, and it should ideally be data on students from earlier cohorts. Lastly, I chose 2014 as it is the last year before the Vacancy Based Dimensioning reform was implemented in 2015, and this allows me to relate the policy experiments to the reform.

The application data does not contain information on the capacities of programs in a given year. The information on program capacities is only publicly available for full programs. I was generously allowed access to this information from the Ministry of Higher Education and Science. The program capacities data contains information, for each program and year, on the capacities by to the different enrolment channels and the number of filled seats. Some programs have unrealistically high capacities, e.g., 999, and these should be interpreted as being higher than the number of filled seats, although the exact number is unknown. As I need these for my belief estimation and policy experiments, I need a more realistic measure of the capacity for these programs. My solution is to use the highest observed capacity for a given program across all available years. I will refine the solution to this problem in a future version of the paper.

Finally, I get information on program-specific requirements from the historic executive orders concerning admission requirements to the higher education programs in Denmark.[6] The program-specific requirements are admission requirements that are additional to the regular admission requirements and vary by program, and I have collected them specifically for this project. In particular, the program-specific requirements data contains requirements such as minimum GPA, minimum requirements for grades in cer-

[5]Applicants in Denmark can multiply their GPA up before they apply by some multipliers based on, e.g., the number of A level courses in high school and if the applicant graduated from high school no more than two years prior to the application.

[6]The historic and current executive orders can be found on www.retsinformation.dk.

tain high school courses, and minimum requirements for the level of certain high school courses.

Based on the application data, I create one of the program-specific characteristics I use to estimate applicants' preferences. The measure is the difference between an applicant's GPA in 2014 and the average GPA of accepted applicants in a given program in 2013, scaled by the standard deviation of the GPA of accepted applicants in the same program in 2013. I calculate the measure as

$$G_{ij} = \frac{GPA_{i,2014} - \overline{GPA}_{j,2013}}{\sigma_{GPA_{j2013}}}$$

where GPA_i is applicant GPA, $\overline{GPA}_{j,2013}$ is the average GPA of accepted applicants for program j in 2013, $\sigma_{GPA_{j2013}}$ is the standard deviation of the GPA of accepted applicants for program j in 2013, and G_{ij} is the standardized GPA of applicant i.

1.3.2 Administrative Register Data and distance measure

I mainly rely on administrative register data to generate the program-specific characteristics. However, to link the program characteristics, generated based on the registry data, with the application data, I first need to generate links between the program identifier in the application data and the program identifier in the education registry. To generate the links, I rely on the student registry (KOTRE), which contains all educational spells. I outline some issues related to forming the links and the approach I use to overcome these in appendix 1.A.

After I have generated the links between the program identifiers in the application data to the program identifiers in the registries, I can generate three additional program characteristics for the expected labor market conditions after the applicants have graduated to help me estimate applicants' preferences.

To generate my three measures of labor market conditions, expected unemployment, expected earnings, and dispersion of expected earnings, I first combine the spell data from the student registry (KOTRE) with the link between program identifiers in the application data and the student registry, to identify all students who enrolled in a relevant bachelors degree in 2002 or later. I then condition on graduating with a bachelor's degree and enrolling and graduating with a subsequent master's degree. I keep the graduation date, month, and year, along with the program link and the personal identifier for students who graduated with their master's degree between 2008-2016. I then combine this data with the monthly employer-employee registry (BFL) for 2008-2016 to generate the measures for expected labor market conditions. To get a measure of expected unemployment, I first calculate the number of days between the date of graduation and the first day in the employer-employee registry, where I record a student as being employed in a position equivalent to 75% of full-time. I then estimate expected unemployment,

U, as the average number of days between graduation and the first recorded day of employment divided by 30 to get months. I rely on a similar approach to get my measures of expected earnings and the dispersion of expected earnings. Instead of using the graduation date, I use the month. I then deflate the monthly earnings during the 12 months after graduation and divide by 10.000. Then I estimate the mean monthly earnings to use as my measure of expected earnings, $\overline{w}$. Finally, I estimate the dispersion of the expected earnings, σ_w as the standard deviation of the monthly earnings.

The distance measure, D, is created for this project by Denmark Statistics. It measures the road distance from the addresses of all high schools in Denmark to the main campus of all universities, such that the measure varies both by program and individual.

1.3.3 Sample selection

The full dataset contains information on $91,276$ applicants who apply to 897 different programs. After initial cleaning of the data, which mostly consists of removing canceled applications and applicants with multiple accepts, missing personal identifiers, and so on, results in a removal of $14,741$ applicants, leaving me with $80,112$ applicants, which I refer to as the *universe of applicants*. I use this sample to describe the overall characteristics of all applicants.

I select two samples from this *universe of applicants* to use in my estimation. I refer to them as the *full sample* and the *analysis sample*, which is a subset of the *full sample*. I use the *full sample* every time I run the allocation algorithm, which is a part of the estimation of subjective beliefs and the policy experiments. I use the *analysis sample* to estimate preferences and to summarize the results for the policy experiments.

To get the *full sample*, I remove all applicants with a missing high school GPA and applications to programs with zero capacities. This selection reduces the sample to $65,214$ applicants who apply to 693 different programs. As mentioned, I use this sample to estimate beliefs. The applicants with missing high school GPAs and applications to programs with zero capacities strictly apply through the quota 2 channel. As they are not tried on the quota 1 criteria, removing them does not affect the belief estimation.

To get the *analysis sample*, I select only applicants with applications strictly through the quota 1 channel, consisting only of university bachelor degrees. Further, I restrict the sample to applicants with an observable high school degree from a Danish high school. Lastly, I restrict the sample to applicants who only include programs with an observable value for all program and program/applicant characteristics used in the estimation (D, G, U, $\overline{w}$, and σ_w). This results in a sample of $12,964$ applicants with $24,025$ applications applying to 234 different programs. The two first selection requirements (quota 1 and bachelor) are highly correlated and the most important in terms of the reduction in sample size both by themselves and together.

1.4 Descriptives

In this section, I describe the overall characteristics of the applicants in the three different samples, the characteristics of the programs I observe in the applications, and the observed application behavior.

1.4.1 Applicant characteristics

Table 1.1 shows overall sample characteristics for three different samples. The table contains three panels, one for each sample. The first column shows mean characteristics, the second column shows standard deviations, and the third column shows the sample percentage with a non-missing value for the variable. By comparing panel C, the *analysis sample*, with the other two panels, we see that the applicants in the *analysis sample*, on average, have a higher GPA, are younger, submit applications with fewer programs, are more likely to be accepted to any program, and are accepted on a higher ranked program in their application conditional on being accepted. Further, we also see that the analysis sample only contains applicants whose applications consist of university programs and only apply through the quota 1 channel, as I condition on these in the sample selection. Programs evaluate applicants through the Quota 2 channel based on measures other than high school GPA, e.g., grades or levels from specific courses in high school, motivational letters, and relevant past work experience, and I, therefore, expect them to have a lower GPA and be older on average.

1.4.2 Program characteristics

Tables 1.2 and 1.3 show the characteristics of the programs in the observed applications for ranks one to three. Since the number of students with longer applications is low, I only rely on the three first ranks to generate moments for the estimation. Table 1.2 shows the mean and standard deviation by rank 1 to 3 of the observed applications for the different program characteristics. Higher ranked programs are, on average, a shorter distance to applicants' high school, while all the other characteristics are evenly distributed across the top three ranks on average. Appendix table 1.D.1 shows mean characteristics for ranks four to eight. The pattern in the distance continues to the lower ranked programs. At the same time, programs ranked lower in the application than rank 3 have higher expected unemployment, lead to lower expected earnings, a lower dispersion in expected earnings, and also have students in the prior application year with lower GPAs.

Table 1.3 shows the share of applicants who applied for a program in a given field[7] on the upper panel and a given university on the lower panel by ranks 1 to 3. Most applicants apply to a program in humanities as their top rank, with programs in social

[7]Tabel 1.B.1 gives examples of which programs belong to the fields in table 1.3

Table 1.1: Summary statistics by sample

Panel A: Universe of applicants	Mean	Std	Share in sample
GPA	7.21	2.53	85.10%
Female	0.53	0.50	100%
Age	22.54	5.62	91.58%
No. priorities	2.40	1.71	100%
Accepted	0.80	0.40	100%
Accepted priority	1.27	0.77	80.33%
Contains university program	0.52	0.50	100%
Only university programs	0.41	0.49	100%
Only through Quota 1	0.45	0.50	100%
Contains Quota 2	0.48	0.50	100%
Observations	80,112		
Panel B: Full sample	Mean	Std	Share in sample
GPA	7.29	2.52	100%
Female	0.55	0.50	100%
Age	21.84	4.85	93.57%
No. priorities	2.50	1.73	100%
Accepted	0.78	0.41	100%
Accepted priority	1.31	0.82	78.08%
Contains university program	0.63	0.48	100%
Only university programs	0.51	0.50	100%
Only through Quota 1	0.57	0.50	100%
Contains Quota 2	0.43	0.50	100%
Observations	65,214		
Panel C: Analysis sample	Mean	Std	Share in sample
GPA	9.10	2.26	100%
Female	0.51	0.50	100%
Age	20.21	2.48	98.43%
No. priorities	1.85	1.21	100%
Accepted	0.87	0.33	100%
Accepted priority	1.11	0.45	87.26%
Contains university program	1.00	0.00	100%
Only university programs	1.00	0.00	100%
Only through Quota 1	1.00	0.00	100%
Contains Quota 2	0.00	0.00	100%
Observations	12,964		

Note: The table contains three panels, panel A is for the *universe of applicants*, panel B is for the *full sample*, and panel C is for the *analysis sample*. The first column contains means for the selected variables, the second contains standard deviations, and the third contains the share of the given sample with non-missing values for the variable in percentages.

Table 1.2: Characteristics of programs in applications by rank

	Rank 1 Mean (Std)	Rank 2 Mean (Std)	Rank 3 Mean (Std)
Distance (10 Km)	7.15	8.45	8.98
	(7.80)	(8.16)	(8.55)
Standardized GPA	-0.19	-0.17	-0.19
	(1.68)	(1.72)	(1.71)
Unemployment (Months)	4.55	4.46	4.55
	(2.70)	(2.66)	(2.77)
Expected earnings (10,000 DKK)	2.50	2.52	2.52
	(0.58)	(0.59)	(0.59)
Dispersion of Expected earnings	1.14	1.18	1.16
	(0.44)	(0.57)	(0.52)
Observations	12,964	5,965	2,956

Note: The reported numbers are means (standard deviations in parentheses). The first column reports variable names and units in parenthesis. The columns indicate for which rank in the applications the measures are. The number of observations shows the number of applicants with at least the given number of ranks in their application.

sciences (including economics and political science) in a close second place, followed by business (including other business and law). The patterns look similar for ranks 2 and 3 when conditioning on having at least two or three programs in the application. The pattern for universities is also quite similar across the top 3 ranks. Most applicants, 37%, have a program at the University of Copenhagen as their top rank, and the second most popular university is Aarhus University, with 20% of applicants choosing a program there as their top rank.

1.4.3 Application behaviour

Figure 4 shows applicants' overall application patterns in the *full sample* and the *analysis sample*. The upper panel shows the distribution of the length of the applications. Most applicants submit a short application with only one program, 54% of the *analysis sample*. The overall pattern is that as the length of applications increases, the share of applicants who have submitted an application of the given length falls. This pattern is clearer for the *analysis sample* compared to the *full sample*, with almost none of the applicants in the analysis sample submitting applications with five or more programs. The lower panel shows the unconditional distribution for on which rank applicants are accepted. We see that most applicants are accepted on their top ranked program and that this is more pronounced for the *analysis sample* (80.4%) compared to the *full sample* (63.9%). Further, we see that 12.7% of the *analysis sample* are rejected from all programs in

Table 1.3: The distribution of programs in different fields and universities by rank

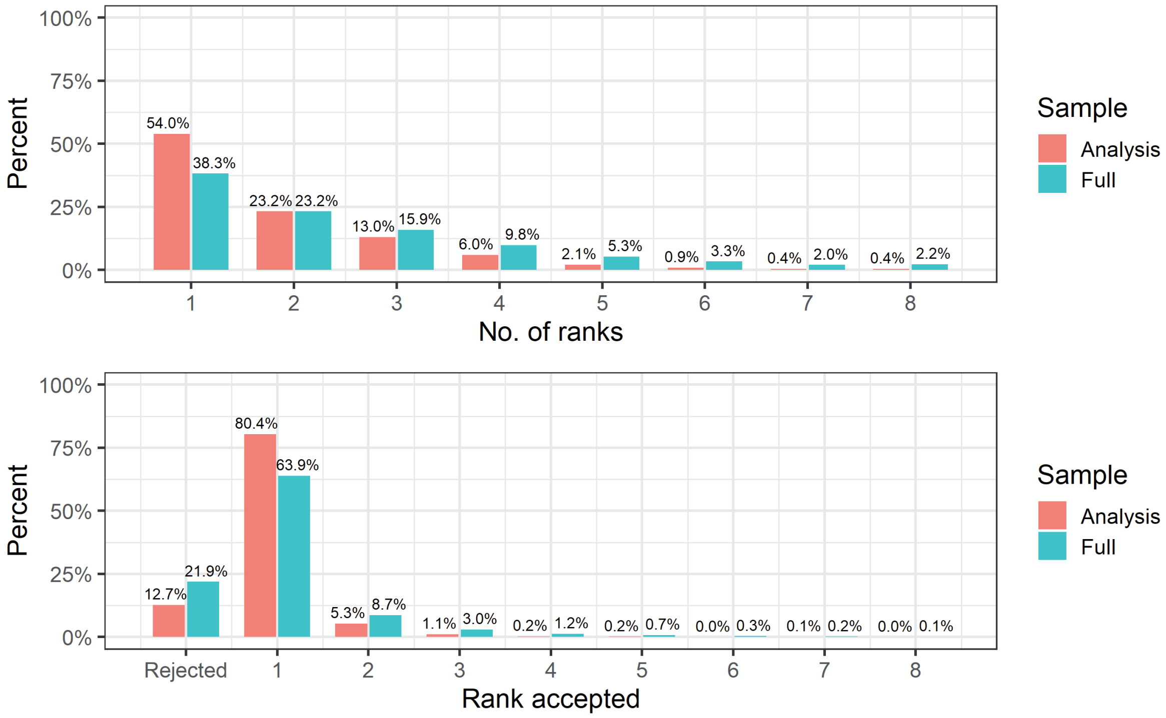
Field	Rank 1	Rank 2	Rank 3
Social science	0.12	0.13	0.14
Humanities	0.20	0.21	0.21
Health	0.05	0.03	0.03
Natural science	0.21	0.22	0.22
Engineering	0.06	0.04	0.04
Other Business	0.08	0.10	0.09
Education	0.01	0.01	0.01
Economics	0.03	0.03	0.03
Medicine	0.08	0.08	0.09
Law	0.08	0.07	0.08
Political science	0.03	0.03	0.03
Business	0.06	0.05	0.05
University	Rank 1	Rank 2	Rank 3
University of Copenhagen (KU)	0.37	0.33	0.36
Aarhus University (AU)	0.20	0.21	0.21
Copenhagen Business School (CBS)	0.08	0.11	0.11
Aalborg University (AAU)	0.12	0.09	0.08
University of Southern Denmark (SDU)	0.13	0.15	0.13
Roskilde University (RUC)	0.05	0.06	0.05
Technical University of Denmark (DTU)	0.05	0.05	0.05
IT University of Copenhagen (ITU)	0.01	0.01	0.01
Observations	12,964	5,965	2,956

Note: The table shows the share of applicants who have ranked a program in a given field (top panel) or university (bottom panel) for the first three ranks in their applications. The number of observations shows the number of applicants with at least the given number of ranks in their application.

their application, while this is 21.9% for the *full sample*. Further, we see that the share of applicants accepted on a given rank is falling with the number of ranks in the applications. The general pattern of short listing combined with the fact that low GPA applicants apply to programs with lower cutoffs the year before is in my interpretation a clear indication of strategic behaviour, where applicants with low GPAs top censor their applications, while applicants with high GPAs only include their most desired programs.

Figure 1 shows the distribution of length of application in the top panel and which rank the applicant was accepted on, if any, on the bottom panel by GPA for the *analysis sample*. In the top panel, we see that applicants are almost equally likely to submit an application of a given length across the GPA distribution. If we turn to the bottom panel of figure 1 we see that applicants are also almost as likely to be accepted on a given

Figure 1.1: Application patterns for the full sample and the analysis sample

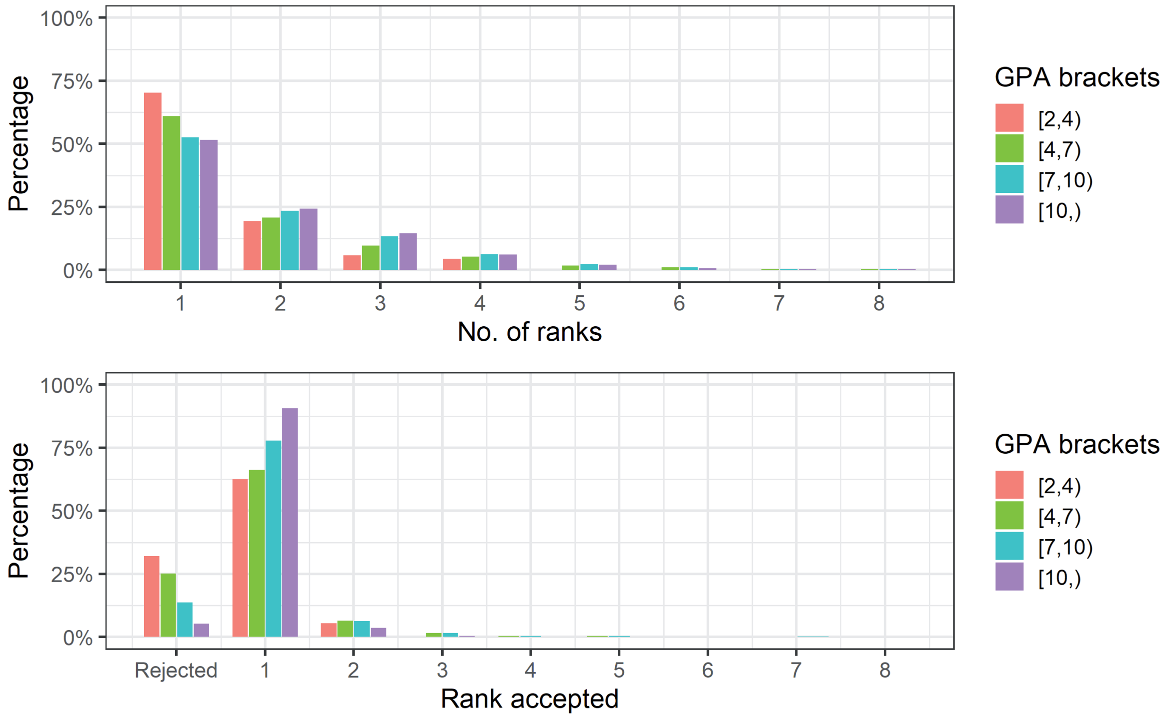


Note: The upper plot shows the share of applicants with a given ROL length, with percent on the second axis and ROL length on the first axis. The lower plot shows the distribution of accepted rank, if any.

priority across the GPA distribution. I interpret this finding as evidence of applicants not reporting their preferences truthfully. Almost no applicants reach the maximum length of their applications, so including one or more programs with historically higher GPA cutoffs on a higher rank would not affect their probability of being accepted in their current ranked programs.

Figure 7 shows the variation in cutoffs from 2013 to 2014. The first thing to notice is the two lines that the points sketch out. It is only feasible for points to be located on the lines or below the diagonal line and on the right of the vertical line. Programs on the bottom vertical line have the lowest possible cutoff in 2014, a GPA of 2, which means that the change from 2013 to 2014 can only be negative or zeros. Points on the top diagonal line had the lowest possible cutoff in 2013, a GPA of 2, so the resulting change will be equal to the cutoff in 2014 minus 2. The main takeaways from the figure are that there is much variation in yearly cutoffs and that many programs have non-binding cutoffs (although most are non-university programs). So applicants cannot predict the cutoffs perfectly. Remember, capacities and capacity changes are not public, and further students might not be able to predict how many other applicants there are or what their GPAs are.

Figure 1.2: Application patterns for the full sample and the analysis sample by GPA



Note: The upper plot shows the distribution of the length of applications by GPA groups. The number of ranks in the applications is on the first axis, and shares in percentages are on the second axis. The lower plot shows the distribution for which rank an applicant is accepted on or if the applicant is rejected by GPA groups. For both plots, applicants are grouped by GPA accordingly 2 to 4, not including 4, 4 to 7, not including 7, 7 to 10, not including 10, and 10 or above.

1.5 Model

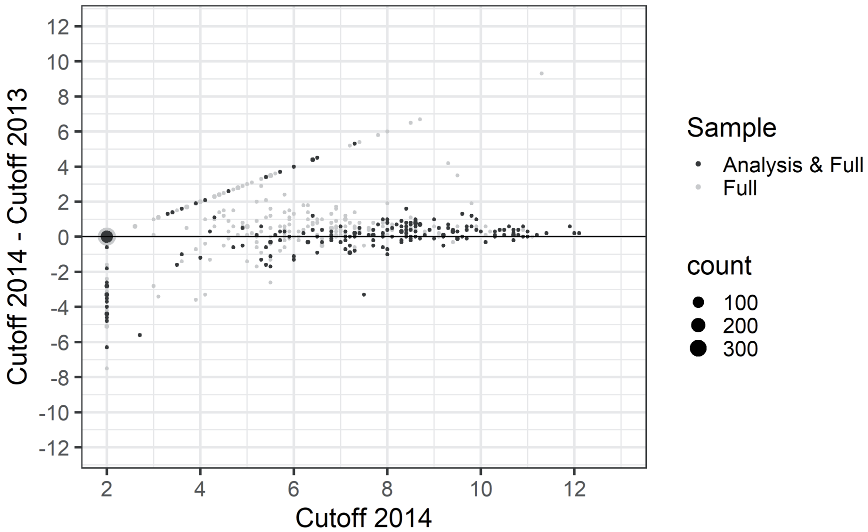
In this section, I formalize the applicants' education choice problem. In essence, the problem for the applicants is to optimally choose to submit a ROL that maximizes their expected utility. I base the framework on the optimal portfolio choice model in Chade and Smith (2006).

There are $i = 1,...,N$ applicants choosing up to $\bar{R}$ out of J programs to include in their ROL, which they submit to the CAS. I denote the set holding all possible ROLs, e.g., all combinations of programs for any given length of ROL up to $\bar{R}$, $\mathcal{R}$, and individual i's chosen ROL from all of the possible ROLs as R_i. The length of R_i, $|R_i|$, gives the number of ranked programs in R_i. All programs have positive capacities, $q_j > 0\,\forall j$. Further, all programs have identical and known preferences over applicants' scores.

All applicants have preferences over all j programs, sorted according to their indirect utilities u_{ij}, and all applicants have subjective beliefs about their admission chance to all j programs expressed by the vector given by $p_{ij} \in [0,1]$. I further assume beliefs are independent across programs.

I can then formalize applicant i's expected utility from submitting a given ROL R_i

Figure 1.3: Change in cutoffs from 2013 to 2014



Note: The figure shows the change in cutoffs from 2013 to 2014. The first axis shows the cutoff in 2014, and the second axis shows the difference from 2013 to 2014. The number of programs scales the size of the points at a given point. The color of the points indicates whether a given program is only part of the *full sample* (grey) or part of the *analysis sample* as well (black).

as

$$\mathbb{E}U(R_i) = \sum_{k=1}^{|R_i|}\left(\prod_{r=1}^{r_k-1}(1-p_{ir})p_{ir_k}u_{ir_k}\right) + \prod_{k=1}^{|R_i|}(1-p_{ik})u_{i0}, \tag{1.1}$$

where u_{ij} is the indirect utility from being accepted to program j, p_{ir} is the individual and program-specific belief about the chance of being accepted to the program ranked as r. The k term in r_k describes the program's rank in the applicant's ROL. The last term is the value of the outside option. Applicants not accepted to any programs included in their ROL receive the value of the outside option, which I set to 0. Further, I require all applicants to list at least one program in their ROL, so my model does not capture the extensive margin fo the application behavior, e.g., whether to apply at all or not. The expected utility reflects the characteristics of the DA assignment mechanism in place, considering that an applicant will only be tried for admission at program j if she is rejected from all programs she ranked higher in her application.

The applicants' problem is then to maximize the following expression

$$\max_{R_i \in \mathcal{R}} \mathbb{E}U(R_i) - c(|R_i|), \tag{1.2}$$

where $\mathbb{E}U(R_i)$ is the expected utility from submitting ROL R_i and $c(|R_i|)$ is the application cost function, which depends on the length of the submitted ROL. As there are no pecuniary application costs, the cost function contains only non-pecuniary application costs, such as the information cost associated with finding programs to include in the application. I assume the cost function is linear $c(|R_i|) = |R_i| \cdot c$, where c is a small and fixed application cost. Further, I let the application cost function go to infinity for $|R_i| > 8$ as applicants cannot submit more than 8 programs in their applications. Finally, I calibrate the fixed cost c to $1e-6$ to ensure that applicants only include programs in their application that strictly improve the expected utility. Otherwise, it is always optimal for applicants to include programs with a subjective probability of 0 in their applications as long as they have ranked fewer than eight programs.

Applicants then sequentially add programs, which increase the expected utility from submitting a given ROL, taking into account the cost of submitting the ROL. Applicants can either include programs in the top, the middle, or the bottom of their ROL, and Haeringer and Klijn (2009) show that it is optimal for applicants to sort the programs in their chosen ROL by expost utilities.

1.5.1 Indirect utility

The indirect utility applicant i receives from being accepted to program j is given by

$$u_{ij} = \alpha_j^F + \alpha_j^U + Z_j^P \alpha + Z_{ij}^S \beta + \varepsilon_{ij} \tag{1.3}$$

where α_j^F and α_j^U are field and university fixed effects, Z_j^P is a matrix containing program-specific characteristics given by

$$z_j^P = U_j \alpha_1 + \overline{w}_j \alpha_2 + \sigma_j^w \alpha_3, \tag{1.4}$$

where U_j is the expected unemployment, $\overline{w}_j$ is the expected earnings, and σ_j^w is the dispersion of expected earnings. Further, Z_{ij}^S is a matrix containing characteristics that vary across applicants and programs

$$z_{ij}^S = d_{ij}\beta_1 + G_{ij}\beta_2, \tag{1.5}$$

where d_{ij} is distance and G_{ij} is standardized GPA. Lastly, ε_{ij} is an additive idiosyncratic taste shock following a type I extreme value distribution.

1.5.2 Solving Portfolio Choice model

Solving the Portfolio Choice model requires finding the optimal portfolio from all possible portfolios for a given applicant. As the number of programs is $J = 234$ and applicants can rank up to $\bar{R} = 8$ programs in their application, this leads to $\binom{234}{8}$ possible combinations, which have to be evaluated for each applicant. Comparing all of them is infeasible. Instead, I use the Marginal Improvement Algorithm (MIA) proposed by Chade and Smith (2006) to find the optimal ROL for a given applicant. Chade and Smith (2006) show that if beliefs are independent and the cost function only depends on the length of the portfolio, it is only necessary to evaluate up to $\frac{234(234+1)}{2}$ portfolios and in most instances much fewer. The algorithm runs in the following steps

1 Start with $R_i^0 = \emptyset$

2 Choose $j = \max_{j \in J \setminus R_i^{n-1}} \mathbb{E}U(R_i^{n-1} \cup j)$

3 Set $R_i^n = R_i^{n-1} \cup j$ and order R_i by u_{ij}

4 Stop if $U(R_i^n) - U(R_i^{n-1}) < c(|R_i^n|) - c(|R_i^{n-1}|) = 0$ or all remaining $u_{ij} < u_{i0}$

In other words, the algorithm takes a given ROL, R_i, and calculates the expected value from including the remaining non-chosen alternatives in J one at a time. It thereafter checks which alternatives improve the expected value of submitting the ROL and chooses the one that gives the highest improvement in the expected value, if any. In other words, as the name suggests, the algorithm looks for the highest marginal improvement to the expected portfolio value, if any exists.

1.6 Identification

Identifying preferences for college programs is a challenging task. It is possible to rationalize every possible ordering by only looking at applicants' submitted ROLs, as it is impossible to separate strong preferences on unobservables from, e.g., beliefs. To identify applicants' preferences and beliefs separately, I rely on including what Agarwal and Somaini (2018) call a special regressor along with an assumption of rational expectations. The purpose of the special regressor is to include an exogenous variable, which only shifts preferences through indirect utility and not through beliefs. As Agarwal and Somaini (2018) suggest I include an individual-level measure of distance to education programs which I assume is exogenously determined. The specific distance measure I use is the road distance between applicant is high school and the location of program j, so my distance measure only varies for programs across universities and not within universities. This means that I also rely on the other measures in the indirect utility function, which vary across programs within universities as well, to separately identify preferences from beliefs.

Manski (2004) suggests that researchers circumvent the difficulties of separately identifying preferences and beliefs by eliciting subjective beliefs with, e.g., a survey. The approach suggested by Manski (2004) is not feasible in my case for two reasons. Firstly I rely on data on observed preferences, which was already collected, so I cannot go back and survey applicants. Secondly, my counterfactuals require me to be able to model applicants' belief formation. My approach is instead, to assume that applicants have rational expectations when they form their beliefs and that the beliefs can be estimated using the bootstrap estimator proposed in Agarwal and Somaini (2018).

To further help me identify the parameters of the indirect utility function, I use the variation in program characteristics and individual-program program characteristics in my data. I describe the different sample moments I am exploiting to identify the different parameters in section 1.7.

Changes in capacities are not announced prior to applications, and students can only infer changes to capacities for programs with a binding cutoff expost, conditional on the fact that the program had a binding cutoff in the prior year as well. I assume that changes to capacities happen exogenously from the student's perspective and that students do not base their application decisions on expectations for how program capacities might change from the previous year.

1.7 Estimation

In this section, I describe the estimation procedures used for the Portfolio Choice model and applicant beliefs.

1.7.1 Portfolio Choice model

The estimation of the portfolio choice model is not straightforward. As I mentioned in section 1.5, there are many possible portfolios and it is not feasible to solve for them all. Finding an expression for the choice probabilities would require me to find the expected value for all possible portfolios. As there is no convenient expression for the choice probabilities, I cannot use Maximum Likelihood estimation. What I do instead is to solve the model using MIA for some given applicant and program characteristics along with simulated taste shocks. I can then use the Simulated Method of Moments (SMM) for estimation. The SMM method implies simulating applications for a given guess on parameters and comparing the chosen moments from the simulated data with the same moments from the observed data and, in essence, minimizing the weighted distance between these. The SMM objective function is given by

$$Q(\theta) = (M_{simulated}(\theta) - M_{data})'\Omega^{-1}(M_{simulated}(\theta) - M_{data}), \tag{1.6}$$

where M_{data} is a vector of moment conditions from the data and $M_{simulated}(\cdot)$ is the corresponding moments based on simulated data from the model for a given guess on the parameters θ. Ω is a weighting matrix. The initial estimation uses the identity matrix as the weighting matrix. Although the estimator is consistent with large N and a fixed number of simulations S, it will most likely be inefficient as it gives equal weight to all moment conditions, even though I do not expect that all moments are equally informative. The current choice of weighting matrix is a result of time constraints, as the model is computationally costly to solve and simulate. Later implementations will use a more efficient weighting matrix.

I select the moments to include based on the features I want the model to capture in the data. I first discretize all the continuous distributions for program characteristics (distance, standardized GPA, expected unemployment, expected earnings, and the dispersion of expected earnings) based on the application data. By using some discretization points I can calculate the empirical and simulated share of applicants within a given group for a given characteristic by, e.g., rank in their application, field, and university. I interpret the calculated shares as probabilities, and the estimation procedure then minimizes the distance between the empirical and simulated probabilities, where I first take the average over the simulations for the simulated moments.

I estimate moments in M_{data} which are based on the continuous measures with the following function

$$M_{kr} = \frac{\sum_i^{N_r} \mathbb{1}_{\{cut_{k-1} < variable_{ir} \geq cut_k\}}}{N_r}$$

where cut_k is the kth discretization point for a given variable, N_r is the number of applicants who have any program ranked as rank r, $\mathbb{1}$ is in indicator function, $variable_{ir}$ is the value of a given program characteristic for the program ranked on rank r by individual i. Hence, M_{kr} is the share of applicants out of the N_r, who have ranked programs on rank r with values of a given program characteristics within the interval between cut_{k-1} to cut_k. Table 1.4 displays my chosen discretization points.

Table 1.4: Discretization points for the moments

Measure	Cutpoint (cut_k)
D	$0 < 5 < 15 < \infty$
G	$-\infty < 0 < \infty$
U	$0 < 6 < 12 < \infty$
$\overline{w}$	$0 < 2.5 < 3.5 < \infty$
σ_w	$0 < 1 < 1.5 < \infty$

Note: The table reports the discretization points used to generate the moments for the simulated method of moments estimation.

In the estimation I use the discretized distributions of the measures directly and also include the share of applicants with any program on a given rank, as well as the share of applicants who rank a program in a given field or university and moments based on discretized measures of the continuous variables, which I have interacted with field and university dummies. I further only use the mentioned moments for ranks 1-3 and only include moments that are non zero for the observed data. This gives me a total of 781 moments.

In practice I find the reported parameter estimates in two steps. First I hand calibrate the parameters to values that seem close to a minimum for the criterion function, thereafter I feed these calibrated parameters as initial values to the 'fminsearch' minimizer in Matlab which uses the Nelder-Mead simplex algorithm to estimate the parameters.

I have not estimated standard errors for the estimated parameters in the current estimation framework. Estimating standard errors requires me to estimate the variance-covariance matrix for the moments, which is currently too computationally costly to implement for this paper version. I will estimate standard errors in a future version of the paper.

1.7.2 Beliefs

I do not have information on subjective beliefs, but I can estimate them. Agarwal and Somaini (2018) show that under an assumption of rational expectations, it is possible to estimate subjective beliefs consistently using their bootstrap estimator. Their proposed method samples applicants and their applications with replacement from the population of applicants and runs the assignment algorithm to get the cutoffs for each sample. Repeating this many times in a bootstrap routine makes it possible to characterize the distribution of cutoffs. To estimate the individual subjective belief for a program, I calculate the fraction of times individual i's GPA is equal to or above the simulated cutoff for program j. The method requires information on all applicants and their applications and admission scores, as well as information on the mechanism used to match applicants with programs and the capacities of the programs. The bootstrap estimator is given by the following expression

$$\hat{p}_{ij} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{\{s_i \geq P_{jb}\}} \tag{1.7}$$

where $\hat{p}_{ij}$ is the estimated beliefs, $\mathbb{1}$ is an indicator function that is equal to one when the students score, s_i, is greater than or equal to the simulated cutoff. P_{jb} is the simulated cutoff for program j in bootstrap simulation b. I only estimate beliefs for the *analysis sample*, although I need the *full sample* to simulate the cutoff distributions for the programs. The estimation results in a N-by-J matrix which contains the estimated beliefs, $\hat{p}_{ij}$, in each cell.

As mentioned in section 1.4 I draw from the pool of all applications to simulate the different cutoffs, while I only use the *analysis sample* to estimate the beliefs. The reason for doing this is that it is impossible to subset the capacities in a meaningful way, and simulating the cutoffs on a sub-sample of the pool of applicants would result in many simulations where many programs never reach their capacities.

As mentioned in section 2.2, the Danish assignment procedure differs slightly from the standard DA mechanism (mainly as it includes different quotas, standby applications, and applicants cannot submit more than 8 programs). However, knowledge about the allocation mechanism allows me to consider these to get consistent estimates of the beliefs.

Firstly the option to apply for quota 1 standby is trivial to incorporate as it follows the same GPA ranking as quota 1. It is thereby possible to allow for additional standby seats, which only students applying on standby can fill, resulting in additional standby cutoffs for each program with this application channel. It is more difficult to incorporate the quota 2 and quota 2 standby channels. This is because the assignment criteria are unknown/nontransparent and program-specific. The different programs rank applicants based on some point system, which partly relies on subjective evaluations and changes from program to program. This is because the CAS needs a ranking of the applicants to allocate them. Unfortunately, I do not have access to data on these rankings.[8] If I choose not to include quota 2 applicants in my analysis, I can take advantage of the fact that all quota 2 applications are first tested using quota 1 criteria. This implies that $p_{q1} \leq p_{q2}$ or the individual probability of acceptance for quota 2 applicants has a lower bound, which is the individual probability of acceptance through the quota 1 channel. Further, as I am only interested in estimating preferences for quota 1 applicants, I can use the knowledge of whether quota 2 applicants were accepted for a given program through the quota 2 channel as a measure of the quality of their quota 2 application. What I do in practice is that I observe for which programs applicants are accepted on quota 2. If an applicant reaches such a program in the mechanism and is rejected on the quota 1 criteria, I accept the applicant in quota 2.

This approach to estimating beliefs is crucial for conducting my policy experiment, as it requires explicit modeling of the belief formation of applicants, as this is the only channel through which capacities can affect application behavior.

1.8 Results

In this section, I describe the results. I first describe the estimated utility parameters and university and field fixed effects before I validate how well the model fits the observed data.

1.8.1 Estimated preference parameters

Table 1.5 shows the estimated utility parameters for the Portfolio Choice model. Applicants prefer programs close to the high school they attended, where the accepted applicants in the prior year had higher GPAs on average, with lower expected unemployment, higher expected earnings, and a higher dispersion in expected earnings. The estimated coefficient on distance is as expected negative, such that applicants prefer closer programs, proxied by the location of their high school, to programs further away. The negative sign on the coefficient on G means that applicants prefer programs where they have a lower GPA than the average accepted applicant in the previous year. This is in line with a story of applicants not wanting to waste their GPAs and therefore apply to programs with higher GPA requirements. However, the interpretation comes with a limit as applicants are also more likely to rank programs where they have a higher belief

[8] It would likely not help me much, even if I had access to the rankings, as my policy experiments would require me to model the quota 2 evaluation process.

Table 1.5: Estimated parameters

Parameter	Value
D	-0.104
	(.)
G	-0.200
	(.)
U	-0.009
	(.)
$\overline{w}$	0.127
	(.)
σ_w	0.035
	(.)
Q	0.634
Simulations	2
Applicants	12,964
Observations	24,025

Note: The table reports parameter estimates and standard errors in parentheses. The current solution and estimation procedure make the estimation of standard errors too time consuming, so I have not estimated them. The table also contains the value of the minimized criterion function, Q, and the number of applicants and observations used in the estimation.

about being accepted, which requires a relatively high grade compared to the other applicants in the current application year. As the parameter on U is negative, applicants prefer programs with lower expected unemployment. Further, they also prefer programs with higher expected earnings and programs with a higher dispersion of expected earnings. Hence applicants prefer programs with better labor market prospects, e.g., low unemployment and high wages, with a higher chance of high wages.

Tables 1.6 and 1.7 show the estimated fixed effect parameters on field and university fixed effects. The field fixed effects are all in relation to the field *Social sciences, excl. Economics and Political Science* and the university fixed effects are in relation to the *University of Copenhagen*. We see that conditional on the other variables of the utility function, law, business, and medicine are the most preferred fields, followed by other

Table 1.6: Estimated field fixed effect parameters

Parameter	Value
Social science	ref.
	(.)
Humanities	-0.132
	(.)
Health	0.652
	(.)
Natural science	-0.126
	(.)
Engineering	-0.211
	(.)
Other Business	0.618
	(.)
Education	-0.051
	(.)
Economics	0.076
	(.)
Medicine	3.539
	(.)
Law	2.306
	(.)
Political science	-0.894
	(.)
Business	0.955
	(.)

Note: The table reports the estimated parameters for the field dummies. The reference category is Social science.

business and health. Looking at the university fixed effects, we see that conditional on the other variables of the utility function, Roskilde university and Aalborg University are slightly more preferred than the University of Copenhagen, while Copenhagen Business School, the Technical University of Denmark, and the IT University of Copenhagen are the least preferred.

Table 1.7: Estimated university fixed effect parameters

Parameter	Value
University of Copenhagen	ref. (.)
Aarhus University	-0.131 (.)
Copenhagen Business School	-0.351 (.)
Aalborg University	0.203 (.)
University of Southern Denmark	-0.060 (.)
Roskilde University	0.115 (.)
Technical University of Denmark	-0.438 (.)
IT University of Copenhagen	-0.964 (.)

Note: The table reports the estimated parameters for the university dummies. The reference category is University of Copenhagen.

1.8.2 Validation of Portfolio Choice model

To validate the Portfolio Choice model, I make two comparisons. Firstly, I check how some of the moments used in the estimation compare with the simulated, and secondly, I check if the simulated data moments are close to hold-out moments. The first check is less demanding than the second check as the estimation procedure explicitly minimizes this difference. I have used the first three ranks of applications to fit the model. While applicants can rank up to eight different programs in their applications and I therefore have potential hold out moments for five ranks, the share of applicants who submit an application of a given length is falling by the length, so I only use the moments for the fourth rank as hold out moments.

Figure 1.4 shows the distribution of the length of application for the observed applications in the *analysis sample* and the simulated applications from the model. We see that the model captures the overall characteristics of the distribution for length of application quite well, although it over-predicts the share of applicants with short applications (only one program) and under-predicts applicants with longer applications (more than one program). The model seems to capture the share of applicants with four programs in their application quite nicely as well.

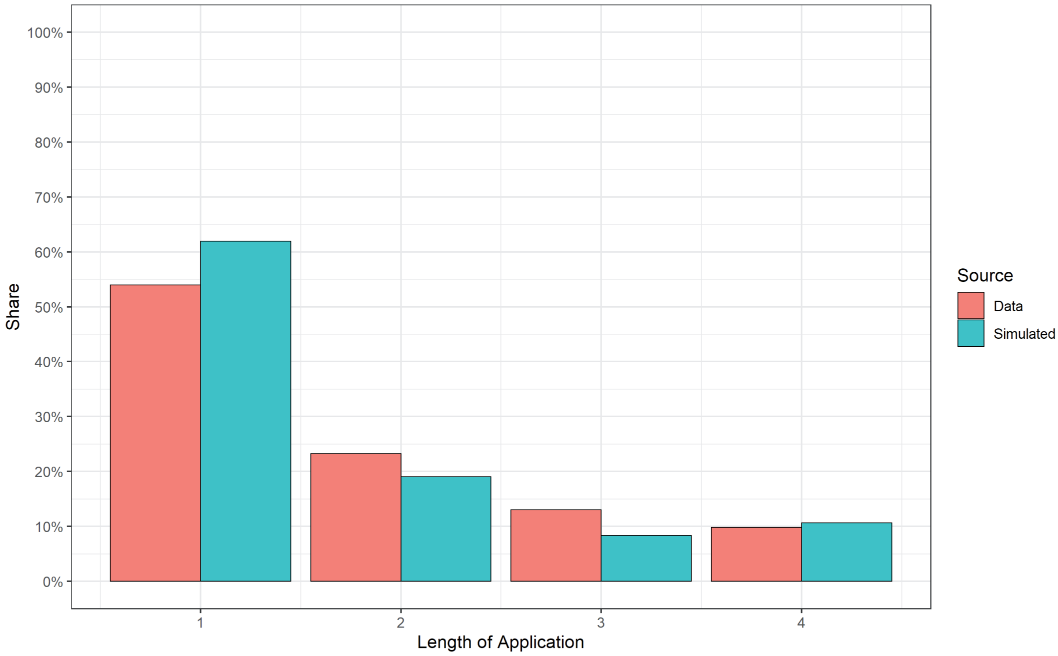
Table 1.8 shows how well the model does in terms of capturing the shares of applicants

Table 1.8: Comparison of data and simulated moments

Panel A: Field	Rank 1		Rank 2		Rank 3		Rank 4	
	Δ	%	Δ	%	Δ	%	Δ	%
Social science	0.039	32.99%	0.024	39.67%	0.011	33.96%	0.000	-3.50%
Humanities	0.005	2.26%	0.005	4.76%	-0.001	-3.03%	-0.007	-31.64%
Health	-0.012	-23.97%	-0.008	-53.33%	-0.003	-48.33%	-0.001	-27.00%
Natural science	0.032	15.29%	0.022	22.17%	0.005	9.49%	-0.008	-37.88%
Engineering	-0.008	-13.61%	-0.011	-62.67%	-0.008	-89.45%	-0.006	-215.79%
Other Business	-0.023	-30.11%	0.006	13.19%	0.000	1.82%	-0.004	-50.00%
Education	-0.008	-100.00%	-0.004	-138.57%	-0.001	-81.82%	-0.001	-90.91%
Economics	0.014	45.10%	0.003	29.00%	0.001	20.51%	0.000	6.41%
Medicine	-0.072	-85.34%	0.017	44.54%	0.018	82.43%	0.010	96.38%
Law	-0.010	-12.60%	0.008	21.86%	0.010	56.09%	0.007	76.23%
Political science	0.021	82.14%	0.012	86.22%	0.006	90.96%	0.002	79.03%
Business	0.021	34.36%	0.004	20.65%	0.001	11.23%	-0.002	-52.50%
Panel B: University	Rank 1		Rank 2		Rank 3		Rank 4	
	Δ	%	Δ	%	Δ	%	Δ	%
KU	0.080	21.8%	0.039	25.7%	0.023	27.8%	0.000	-0.22%
AU	-0.024	-12.3%	0.011	11.8%	0.010	20.4%	-0.003	-16.89%
CBS	0.005	6.7%	0.019	39.0%	0.011	43.1%	0.001	9.50%
AAU	-0.041	-33.2%	-0.015	-33.5%	-0.007	-38.5%	0.002	11.20%
SDU	-0.053	-42.1%	0.003	5.1%	-0.004	-12.7%	-0.007	-52.73%
RUC	0.017	36.3%	0.011	44.0%	0.002	19.4%	-0.001	-29.69%
DTU	0.010	20.5%	0.008	34.5%	0.003	24.3%	0.000	-8.77%
ITU	0.007	62.2%	0.002	47.2%	0.000	23.5%	0.000	27.27%
Panel C: Program char.	Rank 1		Rank 2		Rank 3		Rank 4	
	Δ	%	Δ	%	Δ	%	Δ	%
Distance	-1.054	-14.7%	0.261	3.1%	0.658	7.3%	2.619	25.29%
G	0.003	-1.7%	0.702	-417.5%	1.164	-615.8%	1.265	-350.23%
U	0.332	7.3%	-0.109	-2.4%	-0.118	-2.6%	-0.204	-4.38%
Mearn	-0.070	-2.8%	0.041	1.6%	0.070	2.8%	0.080	3.24%
Stdearn	0.077	6.8%	0.094	8.0%	0.064	5.5%	0.015	1.36%

Note: The table reports differences between empirical and simulated moments (Δ) and the difference in percent of the empirical moment (%) for ranks one to four in the applications. The moments based on the first three ranks are used to fit the model, while the moments based on the fourth rank are hold out moments. Panel A shows how well the model captures the share of applicants who have a program in a given field on ranks one to four. Panel B shows how well the model captures the share of applicants with a program in a given university for ransk one to four. Panel C shows how well the model captures the mean program characteristics for rank one to four.
The field Social science is excluding Economics and Political science and the field Health is excluding Medicine.

Figure 1.4: Empirical and simulated share of applicants by the number of programs in their application



Note: The first axis displays the number of programs in an application (length of the application) and the second axis displays the share of applicants in percentages. The color of the bars indicates the source, red represents the empirical distribution and green indicates the simulated distribution.

who rank a program in a given field (Panel A), shares of applicants who rank a program in a given university (panel B), and the means of program characteristics over the four top ranks. The moments based on the first three ranks are used to fit the model. If we first focus on ranks one to three we see that the model overall captures the moments used in the estimation fairly well. Looking at panel A we see that the difference between the empirical and simulated shares for the different fields is in general below four percentage points ($100 \cdot \Delta$) except for medicine on the top rank, where the model over predicts the share of applicants with a program in Medicine as their top priority by around seven percentage points. The get an understanding of the magnitude of the differences we can look at the columns with the differences in terms of percentages of the empirical share (%). We see that a Δ of 0.072 for Medicine on the top rank corresponds to 85.34% of the empirical share, meaning that the difference is almost as large as the empirical share. If we next move to panel B, we see that the model also does a fairly good job of capturing the shares of applicants with a program in a given university for the first three ranks. The largest difference is that the model under predicts the share of applicants with a program at University of Copenhagen (KU) as their top rank by eight percentage points, although as the share of applicants who rank programs at the University of Copenhagen in general is large, the difference only corresponds to 21.8% of the empirical share. Lastly, if we look at panel C we see that the model seems to capture all the program characteristics except G quite well.

To conclude the model fits estimation moments quite well. The more demanding test is however how the model fits the hold out moments (rank four). For panel A and B the model captures the shares for fields and universities as well and in some instances better than the estimation moments. In panel C the model does almost as good for the hold out moments as the moments used in the estimation, except for distance, where the model under predicts the mean distance to the programs included on the fourth rank by around 26 km ($\Delta \cdot 10$km), which corresponds to 25.29% of the empirical mean distance on the fourth rank.

1.9 Policy experiments

Having described applicant preferences with my model, I now turn to the next point of interest, simulating the effects of supply changes, through changes in capacities, under two different scenarios. Under the first scenario, applicants cannot update their beliefs, and under the second, they can update their beliefs. As the model only captures the demand side of the education market, I can only look at the effects in a partial equilibrium. Therefore, I will not be able to perform a full welfare analysis. I can, however, look at how the different policies affect the distributions of applications, outcomes, and the characteristics of offered programs.

The Portfolio Choice model allows me to perform policy experiments where I change the supply for programs, as the supply only affects the subjective applicant beliefs through the available capacities in the model. Program capacities are only public after all offers are given and only for programs with exhausted capacities. This means that applicants can only see a program's capacity in the previous years (if the capacity was exhausted) and cannot see the change in capacity from the previous year to the current year. While this might give applicants some information on the available capacities, at least for programs with binding cutoffs, it is limited how they can use it when forming their subjective beliefs, as they cannot take possible changes into account.

The section proceeds as follows. I first describe how the policy changes affect applicants in my model and the channel through which applicants can update their beliefs according to the policy changes. After that, I present and discuss the effects of the proposed reduction in capacities for programs within the field of humanities.

1.9.1 Changes to capacities and belief updating

The proposed policy experiments aim to understand how changes to capacities affect applicants' demand for programs. Capacities enter my model through the CAS matching mechanism. The capacities enter the applicants' expected utility from submitting a given portfolio in the subjective program-specific beliefs and further, also affect which program applicants are offered, if any. To evaluate the effect of the proposed policies, I need to know and be able to implement the allocation mechanism. Further, I need to include a channel through which applicants can update their beliefs according to changes in capacities.

I run the policy experiments in the following procedure. I first simulate the model for different capacities under the current framework, where the changes are unanticipated shocks to the capacities, and applicants cannot consider them when forming their beliefs. Second, I look at the same changes to capacities where I instead reveal them to applicants before they form their applications. This allows applicants to update their subjective beliefs according to the new capacities. In my model, program capacities only affect applicants' expected utility from submitting a given application through their subjective beliefs. The indirect utility derived from being admitted to a given program is unaffected. I can therefore find the effect of the proposed policy experiment by only varying capacities and whether I allow applicants to update their subjective beliefs or not while holding the policy invariant utility parameters fixed.

Applicants update their beliefs according to the algorithm in appendix 1.C. The algorithm runs until the cutoff distributions for all programs converge, where each point in the distributions is the equilibrium cutoff from taking a sample with replacements from the *full sample* of applicants and running the allocation mechanism. In each iteration, I reestimate beliefs and solve the portfolio choice model for the given indirect utility parameters and the new estimated beliefs. When the euclidean norm over the vectors of means and standard deviations of the cutoff distributions are below the tolerance parameter ($\epsilon = 1e-6$) I say the algorithm has converged.

Before I run the policy experiments, I need an additional step. The beliefs estimated from the data are based on the observed applications. So to get a baseline for the updating algorithm, which is aligned with the simulations from the model, I reestimate beliefs using the simulated applications from the model. I then use the new beliefs along with the preference parameters in the policy experiments. Figure 1.D.1 illustrates the new predicted shares of applicants with an application containing one to four programs. We see that it has changed very little compared to figure 1.4.

The specific changes I make to capacities are reductions to the capacities in programs within the field of humanities. This is interesting as the vacancy-based dimensioning reform mainly affects programs in the field of humanities.

As the current model does not include an outside option as a choice (I have constrained all applicants to submit at least one program in their applications), a reduction in overall capacities leads to the easily predictive result where fewer applicants are accepted. To avoid this scenario, I instead look at capacity-neutral changes. In other words, when I reduce capacities for some programs, e.g., within humanities, I redistribute the removed capacities among all other programs to keep the same overall capacities. There are many different ways to redistribute capacities, I choose to redistribute capacities according to the distribution of capacities on the program level for unaffected programs, using the baseline capacities. This means that if a program for example had 5% of total capacities excluding programs within humanities before I change capacities, this program will receive 5% of the total capacities that I remove from the programs within humanities.

Lastly, it is important to note, that as I rely on the *full sample* when I run the allocation mechanism, and I only estimate preferences for applicants in the *analysis sample*, the preferences for all other applicants in the *full sample* are fixed when I run the allocation mechanism to look at where applicants are accepted in the policy experiments. All the reported results for the policy experiments only contain the applicants in the *analysis sample*, but their application behavior is also affected by the applications of the other applicants, as the result from the allocation mechanism is a general equilibrium outcome based on the preferences of applicants and programs within the constraint of the capacities. In a future version of the paper I will look more closely at how this affects my results.

1.9.2 Reducing capacities for humanities

Table 1.9: Capacities for the different policies

	$\gamma = 1$	$\gamma = 0.9$	$\gamma = 0.5$	$\gamma = 0.1$
Capacities				
Social science	2,694	2,776	3,099	3,431
Humanities	4,760	4,289	2,401	483
Health	962	992	1,106	1,223
Natural science	4,428	4,562	5,099	5,635
Engineering	1,014	1,047	1,167	1,290
Other Business	2,180	2,244	2,508	2,775
Education	241	249	276	307
Economics	563	579	647	716
Medicine	1,094	1,126	1,259	1,393
Law	1,292	1,330	1,487	1,644
Political science	296	305	340	376
Business	920	948	1,058	1,171
Total excl. Humanities	15,684	16,158	18,041	19,961
Total	20,444	20,447	20,442	20,444

Note: The policy variable γ, which indicates what share of the capacities in the programs within humanities are left, while I redistribute $1-\gamma$ capacities from the programs within humanities across the programs within the other fields. The capacities in the first column, where $\gamma = 1$ are the baseline observed capacities. Due to the rounding of changed capacities to integers on the program level, the sum of capacities for all columns is not the same.

Table 1.9 shows the field level capacities for different values of the policy parameter γ. The capacities are for all programs included in at least one application for the *analysis sample*. The first column shows the baseline total program capacities by field, and the subsequent columns show how the capacities change by fields for different policy parameter values (γ). We see that the humanities programs' capacity drops as the value of γ falls while the program capacities in the other fields increase. This is a feature of the neutral policy design, where reductions in the capacities of programs within one field are redistributed across the programs in the other fields by the baseline program capacities of unaffected programs. We see that the field with the highest number of capacities in the baseline setting is humanities, with the field Natural science and the fields within business and law (Business, Other Business, and Law) in a close second and third place. The chosen redistribution rule means that I redistribute most of the cut capacities within humanities to programs within the mentioned broad fields.

1.9.3 Applications and program characteristics

Table 1.10: Policy changes to capacities: Applications

	Baseline	No updating			Updating		
γ (Fraction of original capacity)	1	0.9	0.5	0.1	0.9	0.5	0.1
Panel A: Humanities							
Length of application	1.98	0.00	0.00	0.00	0.08	0.23	0.07
Accepted	0.84	-0.01	-0.13	-0.49	0.01	-0.04	-0.08
Rank accepted on	1.11	0.03	0.20	0.52	0.01	0.09	0.15
Number of fields	1.54	0.00	0.00	0.00	0.07	0.27	0.23
Same top-ranked program	1.00	0.00	0.00	0.00	-0.08	-0.29	-0.67
Same top-ranked field	1.00	0.00	0.00	0.00	-0.06	-0.25	-0.65
Panel B: All other fields							
Length of application	1.78	0.00	0.00	0.00	-0.02	-0.11	-0.20
Accepted	0.83	0.01	0.03	0.05	0.02	0.03	0.03
Rank accepted on	1.13	-0.01	-0.06	-0.09	-0.03	-0.05	-0.06
Number of fields	1.52	0.00	0.00	0.00	-0.02	-0.06	-0.11
Same top-ranked program	1.00	0.00	0.00	0.00	-0.06	-0.04	-0.06
Same top-ranked field	1.00	0.00	0.00	0.00	-0.05	-0.03	-0.05
Observations	12,964						
Simulations	2						

Note: Column 1 ($\gamma = 1$) contains the baseline application characteristics and outcomes of the matching mechanism. Panel A is for applicants with a program in Humanities as their top rank in the baseline setting, and panel B is for applicants with other programs as their top rank in the baseline setting. The numbers of columns 2 to 7 are all expressed as row-wise deviations from column 1.

Table 1.10 shows the overall counterfactual characteristics of applications split up into two panels, panel A for applicants with a program within humanities as the top rank in their application with the original capacities and panel B for all other applicants for the different values of the policy variable γ and whether applicants can update their beliefs according to a change in program capacities. If applicants cannot update their beliefs, the submitted applications will stay the same, although the realized outcomes from the matching mechanism can change as the capacities change. The first column with $gamma = 1$ is the baseline setting, where the capacities equal the observed capacities. I express all counterfactual results, columns 2-7, as deviations from the results in the baseline setting. I have fixed the applicants for each panel in the table, so even though some applicants, e.g., change their top-ranked program from a program in humanities (with the original capacities) to a program in another field (with changed capacities), they are part of the results in panel A, and likewise for panel B.

I first compare the applications and the outcome of submitted applications for applicants with a program in humanities as their top priority (panel A) to other applicants (panel B) under the baseline setting ($\gamma = 1$). Overall, applicants with humanities programs as their top priority submit longer applications. At the same time, they are as likely to be accepted to a program and, conditional on being accepted, they are accepted on the same rank in their application as the other applicants. The last two categories are not interesting in the baseline setting, as they measure if the applicant has the same top-ranked program or field as in the baseline setting.

Next, I look at columns two to four, which show the deviation from the baseline setting for different policy parameter values under the no updating setting. We see that the length of applications and the shares of applicants with the same top-ranked program or field are the same as in column one. This is because applicants can only change their applications through changes in beliefs, and in the no updating setting, I do not allow applicants to update their beliefs. As expected, for columns 2-4 in panel A, we see that the share of applicants, who are accepted to any program, falls as I reduce the capacities for the programs in humanities, although the share of accepted applicants does not fall one to one with the reduction in capacities. We also see that the applicants are on average accepted at a lower ranked program in their applications. In columns 2-4 in panel B of table 1.10, we see the opposite pattern, although to a much lesser extent. This stems from the fact that the change in capacities is neutral in total capacities. Hence, programs in fields other than humanities get increasingly higher capacities as capacities within humanities are reduced, as we see in table 1.9.

Next, I turn to the setting where applicants can update their beliefs to take the changes to capacities into account, columns five to eight in table 1.10. By comparing panels A and B, we see that applicants with humanities as their top rank in the baseline setting increase the length of their applications, while the applicants with other fields as their top rank in the baseline setting decrease the length. Further, we see that applicants in panel A have a larger probability of being accepted for a small change in capacities ($\gamma = 0.9$), while they have a slightly lower probability of being accepted for larger changes to capacities ($\gamma < 0.9$). On the other hand, applicants in panel B have a slightly higher chance of being accepted, which increases with the change in capacities. For the measure of which rank the applicants are accepted on, we see that applicants in panel A are accepted to a lower ranked program in their applications, while applicants in panel B are accepted to a higher rank in their applications. Last, we see that applicants in panel A increase the number of fields in their applications and are less likely to rank the same program or field as their top rank when capacities in humanities decrease. For applicants in panel B we see that they include fewer fields in their applications as we change the capacities, while they are slightly less likely to report the same program or field as their top rank.

To sum up, the results from table 1.10 show that when I allow applicants to update their beliefs, applicants in panel A can change their applications such that their chance of being accepted is not affected nearly as much, e.g., 76% compared to 35% in the most extreme case with $\gamma = 0.1$ (columns 4 and 7). We also see that the applicants in panel A diversify their applications by including programs in more fields and changing their top ranked program to another field. For applicants in panel B, we see that they become more confident of being accepted to highly ranked programs in their applications, so they reduce the length of their applications. Further, we also see that a few of them change the top-ranked programs and fields in their applications, likely as they have more favorable beliefs about being accepted to other programs.

Table 1.11: Policy changes to capacities: Characteristics of accepted programs

	Baseline	No updating			Updating		
γ (Fraction of original capacity)	1	0.9	0.5	0.1	0.9	0.5	0.1
Panel A: Humanities							
D	8.30	-0.05	-0.08	-0.22	-0.07	-0.09	-0.12
G	0.51	0.04	0.17	-0.20	0.01	0.00	-0.40
U	7.24	-0.13	-0.85	-2.25	-0.21	-1.14	-3.01
$\overline{w}$	1.88	0.02	0.16	0.44	0.04	0.22	0.60
σ_w	1.04	0.00	0.02	0.07	0.01	0.03	0.09
Panel B: All other fields							
D	7.89	0.01	0.01	0.00	-0.01	0.02	0.01
G	0.44	-0.03	-0.13	-0.24	-0.04	-0.14	-0.21
U	3.67	-0.03	-0.12	-0.18	0.00	-0.12	-0.17
$\overline{w}$	2.65	0.01	0.03	0.04	0.00	0.03	0.04
σ_w	1.12	0.00	0.00	0.00	0.00	0.00	0.00
Observations	12,964						
Simulations	2						

Note: Column 1 ($\gamma = 1$) contains the baseline application characteristics and outcomes of the matching mechanism. Panel A is for applicants with a program in Humanities as their top rank in the baseline setting, and panel B is for applicants with other programs as their top rank in the baseline setting. The numbers of columns 2 to 7 are all expressed as row-wise deviations from column 1.

Table 1.11 shows the overall counterfactual characteristics of the programs where applicants are accepted. The table layout is the same as in table 1.10. The table contains two panels, A and B, and the applicants in panels A and B are the same across the columns. Further, columns 2-4 show the program characteristics compared to the baseline for the setting with no updating, and columns 5-7 show the program characteristics compared to the baseline for the setting with updating. I first compare the baseline characteristics of programs applicants are accepted to for the two panels, column 1 in table 1.11. We see that if we compare panels A and B in the baseline setting, applicants with a program in humanities as their top rank (panel A) are accepted into programs that are further away, have peers with lower relative GPAs', can expect around four months of additional unemployment, and almost $10,000$ DKK lower monthly starting wages on average. Further, applicants in panel A also have a slightly lower dispersion in expected starting wages.

In columns 2-4 of table 1.11, the setting with no updating, we see that, as we decrease capacities for programs in humanities, applicants in panel A are accepted to programs further away, with lower expected unemployment and higher expected monthly starting wages. This is somewhat counter intuitive, but the explanation is that the programs in humanities are also the programs with the highest expected unemployment and the lowest expected starting wages on average. So when we reduce capacities for programs in humanities, many of the applicants in panel A are accepted into programs in other fields. We do not see the same pattern for applicants in panel B, where the characteristics of accepted programs do not change much.

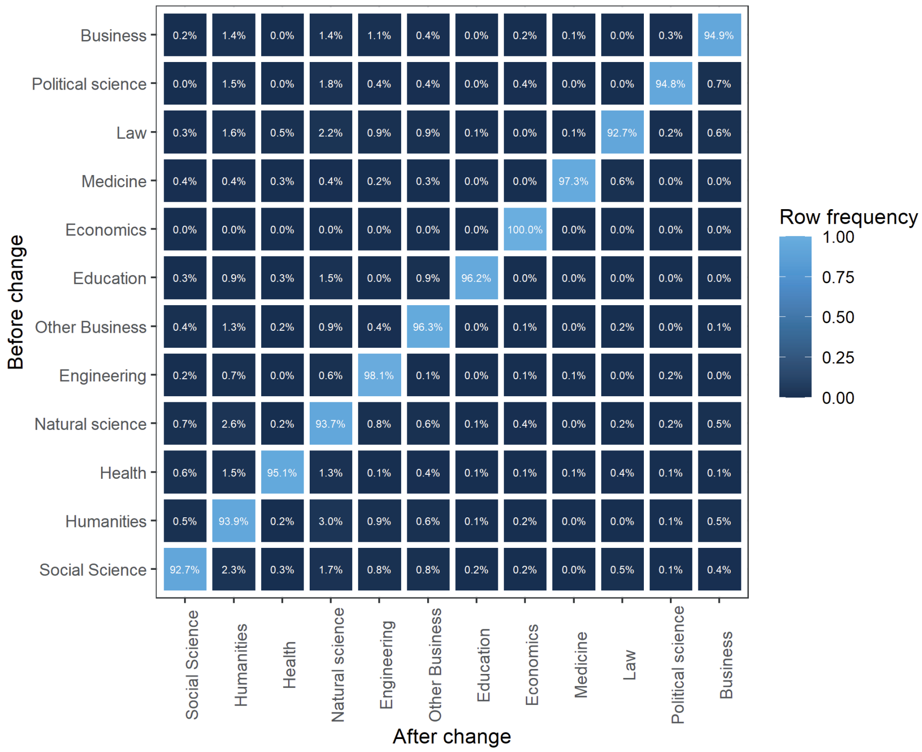
When I instead look at columns 5-7 in table 1.11, which are for the setting where applicants can update their beliefs, I see a similar pattern as in columns 2-4. However, the magnitude of the changes in characteristics is larger for applicants in panel A. In columns 2-4, where applicants cannot update their beliefs, the difference in program characteristics is bounded by the applications in the baseline setting. This is not the case when I allow them to update their beliefs.

So from tables 1.10 and 1.11, I see that when I reduce capacities for programs in humanities while increasing the capacities of other programs correspondingly, applicants who had a program in humanities as their top rank before the change overall see the biggest change. If I do not allow them to update their beliefs in response to the changes, they are less likely to be accepted to a program. However, if they are accepted to a program, the program is closer and has lower expected unemployment and higher expected starting wages. When I, on the other hand, allow them to update their beliefs. In that case, this attenuates the drop in the probability of being accepted, as the applicants can include other programs in their applications, where they are more likely to be accepted. A result of the change in application patterns is that the programs where the applicants in panels A and B in table 1.11 are accepted are much more similar.

1.9.4 Application patterns

Figure 1.5 shows the distribution of top ranked programs by field before (second axis) and after the policy change (first axis) for a 10% reduction in capacities for programs in humanities ($\gamma = 0.9$). The figure illustrates how changes to the capacities affect the

Figure 1.5: Distribution of top-ranked program by field before and after policy change, $\gamma = 0.9$ and applicants update beliefs



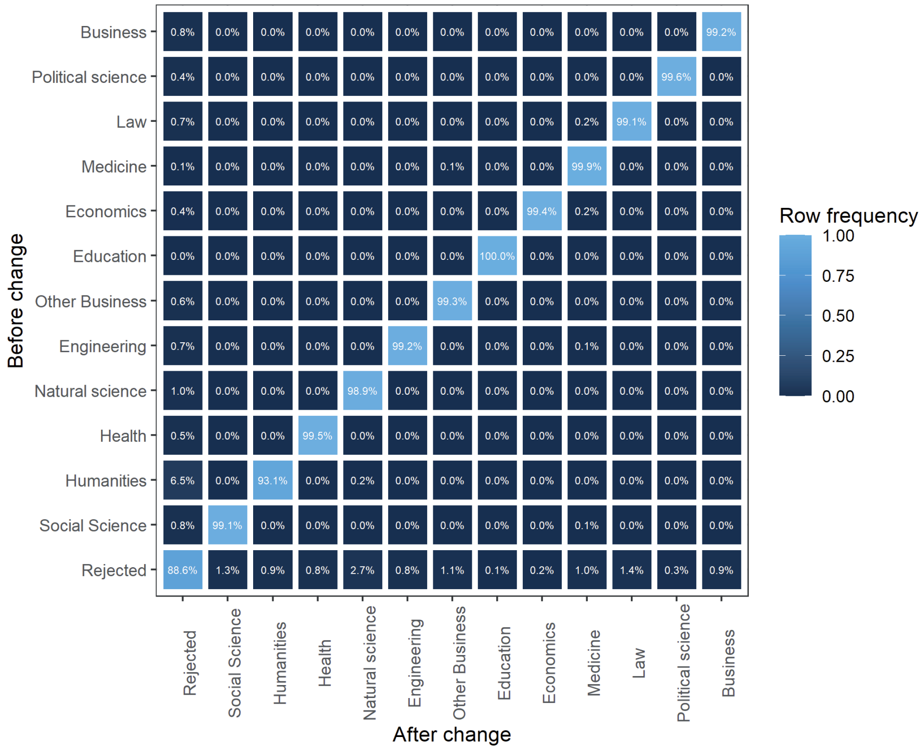
Note: The second axis indicates top ranked field before the change, and the first axis indicates top ranked field after the change. The cells are color coded by the share of applications with a given before/after field combination by field. In other words, the percentages on each row sum to 100%. The policy parameter γ gives the fraction of capacities in Humanities which are left, and conversely, $1-\gamma$ gives the fraction of capacities in Humanities, which are redistributed across the other fields.

application behavior when applicants are allowed to update their beliefs. The case where applicants cannot update their beliefs is trivial, and I do not report it, as applicants do not change their behavior in this case. I first look at applicants with a program in humanities as their top rank (the second to last row) before the change. We see that after the reduction in capacities for the programs in humanities, 93.9% of them still include a program in humanities as their top rank. Interestingly, most applicants either move to natural sciences, social science, or engineering programs. A possible explanation for this pattern is that the applicants who move away from programs in humanities are most likely at the lower end of the ability distribution. As programs in natural sciences typically face low demand and hence have low GPA cutoffs or free slots even applicants with low ability have relatively high beliefs about being accepted in them. Further, as they also lead to expected wages in the top end and a low expected unemployment these programs are attractive to these applicants. If we turn to the other rows of figure 1.5, we see that it is not just applicants who had a program within humanities as their top rank who changed their top ranks. Some of the applicants with programs in other fields also change their top ranks, even though these programs all have increased capacities. This might seem puzzling at first, but as the offers arise in the equilibrium, where the algorithm matches applicants and programs according to their preferences, changes to capacities will affect the application behavior in my model for applicants who are not directly affected by, e.g., a reduction. Further, the figure only reports the top ranked program, and the affected applicants might just have ranked their previous top ranked program lower in their application. Appendix figure 1.D.2 where I have reduced the capacities for programs in humanities by 50% ($\gamma = 0.5$) shows a similar pattern to figure 1.5. However, the share of applicants who had a program in humanities as their top rank before the change is even lower, and the share of applicants with programs in other fields as top rank before the change changes less than in figure 1.5. I can explain the last part by the fact that all other programs now have even more capacities.

1.9.5 Accepted programs

Next, I look at where applicants are accepted. I first look at where applicants are accepted when they cannot update their beliefs, figure 1.6, before I look at where applicants are accepted, when they can update their beliefs, figure 1.7. Figure 1.6 shows the distribution of programs where applicants are accepted by field, before (second axis) and after the policy change (first axis) for a 10% reduction in capacities for programs in humanities ($\gamma = 0.9$) where applicants cannot update their beliefs. Compared to figure 1.5, figure 1.6 has an additional row and column for applicants who are rejected from all programs in their applications. Further, when the applicants cannot update their beliefs, the applications are fixed while the capacities change. Most of the rows in figure 1.6 are not so interesting as nearly all applicants accepted in a program in a given field before the change are also accepted in the same field after the change. The only two rows where something happens are the third last row (humanities before change) and the last row (applicants rejected before change). We see that 6.5% of applicants accepted to a

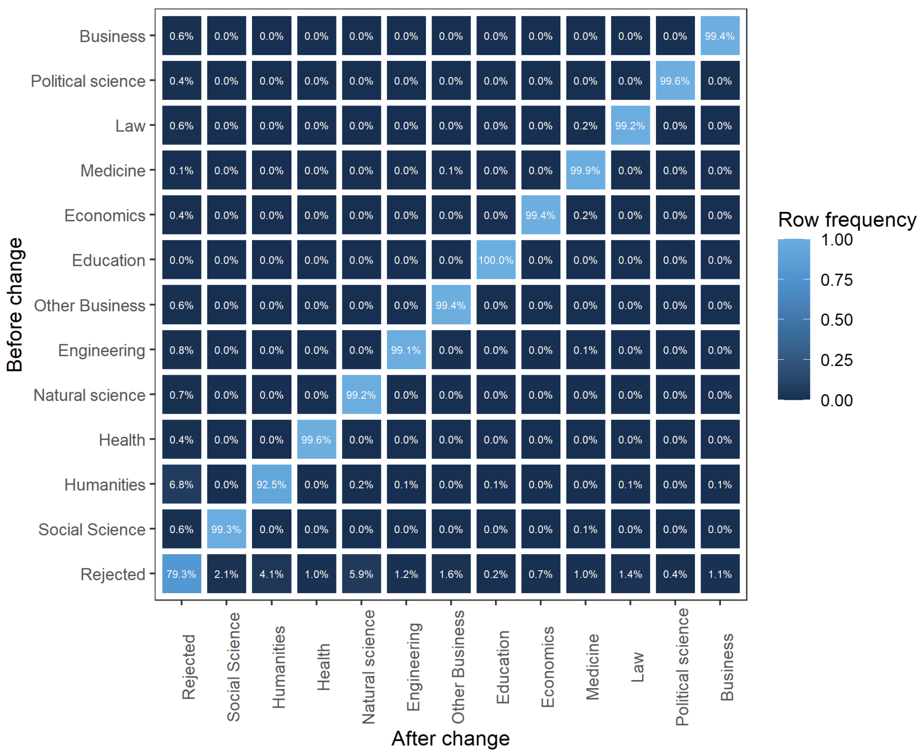
Figure 1.6: Distribution of accepted program by field before and after policy change, $\gamma = 0.9$ and applicants cannot update beliefs



Note: The second axis indicates top ranked field before the change and the first axis indicates top ranked field after the change. The cells are color coded by the share of applications with a given before/after field combination within a before change field, in other words, the percentages on each row sum to 100%. The policy parameter γ gives the fraction of capacities in Humanities which are left, and conversely $1 - \gamma$ gives the fraction of capacities in Humanities, which have been redistributed across the other fields.

program in humanities before the change are rejected, and close to 0% are accepted in programs in other fields. In the last row, we see that the share of applicants who were rejected before the change is reduced by 11.4%.

Figure 1.7: Distribution of accepted program by field before and after policy change, $\gamma = 0.9$ and applicants can update beliefs



Note: The second axis indicates top ranked field before the change, and the first axis indicates top ranked field after the change. The cells are color coded by the share of applications with a given before/after field combination within a before change field. In other words, the percentages on each row sum to 100%. The policy parameter γ gives the fraction of capacities in Humanities which are left, and conversely, $1-\gamma$ gives the fraction of capacities in Humanities, which have been redistributed across the other fields.

Figure 1.7 shows the distribution of programs where applicants are accepted by field, before (second axis) and after the policy change (first axis) for a 10% reduction in capacities for programs in humanities ($\gamma = 0.9$) and applicants can update their beliefs. As figure 1.6, figure 1.7 also has an additional row and column for applicants who are rejected from all programs in their applications. Allowing applicants to update their beliefs according to the change in capacities creates a larger shift in which program applicants are accepted, although as we see in figure 1.7, it is again only applicants who were accepted to a program in humanities or rejected from all programs before the change who are accepted to other programs or rejected. In the third to last row, we see that applicants who were accepted to a program in humanities before the change are now slightly more likely rejected than in figure 1.6. Further, we see that previously rejected applicants in the last row are now more likely to be accepted compared to figure 1.6. The model prediction that more applicants who previously were accepted to a program in humanities are rejected from all programs in the updating setting (figure 1.7) than in the no updating setting (figure 1.6) when I change the capacities is unexpected in the sense that I would expect them to include other programs with a better chance of being accepted (increased capacities). The likely mechanism behind this is that applicants with low GPAs are most likely to change their applications. Since programs in humanities have many applicants with low GPAs, these applicants will also be at the bottom of the distribution within other fields and, therefore, more likely rejected, even though the capacities of other programs have increase. In appendix figure 1.D.4 I reduce the capacities by 50% ($\gamma = 0.5$) instead of 90% in figure 1.7. I see the same pattern in figure 1.D.4; it is mostly applicants who were accepted to programs in humanities or rejected from all programs which are affected by the change in capacities. The effect is larger than in figure 1.7 and not one-to-one with the change in capacities.

To sum up, the reduction to capacities in humanities, and conversely an increase in all other programs' capacities, causes applicants in all groups to change their top ranked fields for small changes in capacities. However, for larger changes in capacities, it is mainly applicants who had a program in humanities as their top rank before the change who are affected, although the change is less than one-to-one with the change in capacities. For small changes, I explain this by excess capacities in programs in humanities. For large changes in capacities, I interpret it as applicants in humanities having strong preferences. Further, we also see that mainly applicants who were accepted to a program in humanities or rejected from all programs before the change are accepted to programs in other fields or rejected after the change. This holds for the setting where applicants cannot update their beliefs and the setting where they can update their beliefs.

1.10 Conclusion

In this paper, I have studied the effect of changes to the supply of higher education programs on the applicant demand. I have in particular studied how a ”neutral” reduction in capacities for programs in humanities affects applicants. The matching outcomes from the mechanism that is used by the CAS to match applicants with programs are general equilibrium outcome, meaning that whether an applicant is accepted or rejected for a given program can affect whether or not other applicants are accepted in other programs. In this sense changes to capacities for some programs can even affect applications to programs with unchanged capacities. I found that applicants in general respond to the change, when they can update their beliefs accordingly. Both the applicants who are directly affected through the reduction in capacities for programs in humanities, and also applicants to other programs which have increased capacities. Further, I found that the response is not one-to-one with the changes to capacities, especially for applicants who applied to programs in humanities before the change, as they have strong preferences for these programs. Lastly I found that making capacities and changes to capacities public before applicants submit their applications can help applicants who otherwise would be rejected from all programs in their applications.

Based on these results I suggest that policy makers are careful when they adjust capacities for higher education programs, as doing so may have unintended effects if it is not done in a careful way.

A straightforward extension to the current model is to extend it to the full universe of applicants. I plan to do this at a later stage and it requires information on the characteristics of the universe of applicants and programs.

Another more challenging extension would be to include the extensive margin of applications to the model as an outside option. This requires including a first step in the model, where applicants choose between the option to apply or not. This allows the study of how the extensive margin of applicants is affected by changes to capacities.

Finally, a very challenging extension to the current model is incorporating the education market's supply side into the model. This requires modeling of the decisions by universities on how to spend their funds. This allows one to look at the full equilibrium effects of changes to capacities and the effect of making universities declare their available capacities publicly before applicants submit their final applications. Such a model would make it possible to understand how not just applicant demand but also universities would respond, e.g., by looking at the trade-off between quality and quantity in teaching and the decision of time allocation of faculty for teaching and research.

References

Abdulkadiroglu, A., & Sönmez, T. (2003). School Choice: A Mechanism Design Approach. *American Economic Review*, *93*(3), 729–747.

Agarwal, N., & Somaini, P. (2018). Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism. *Econometrica*, *86*(2), 391–444. https://doi.org/10.3982/ECTA13615

Agarwal, N., & Somaini, P. (2020). Empirical Analysis of School Assignment Models. *Annual Review of Economics*, *12*(1), 471–501.

Ajayi, K., & Sidibe, M. (2015). An Empirical Analysis of School Choice under Uncertainty. *Working Paper*, 48.

Artemov, G., Che, Y.-K., & He, Y. (2020). Strategic 'Mistakes': Implications for Market Design Research. *Working Paper*.

Azevedo, E. M., & Leshno, J. D. (2016). A Supply and Demand Framework for Two-Sided Matching Markets. *Journal of Political Economy*, 35.

Balinski, M., & Sönmez, T. (1999). A Tale of Two Mechanisms: Student Placement. *Journal of Economic Theory*, *84*(1), 73–94. https://doi.org/10.1006/jeth.1998.2469

Calsamiglia, C., Chao, F., & Güell, M. (2017). Structural Estimation of a Model of School Choices: Boston Versus Its Alternatives. *NBER Working Papers*, (24588).

Chade, H., & Smith, L. (2006). Simultaneous Search. *Econometrica*, *74*(5), 1293–1307.

Chen, Y., & He, Y. (2021). Information acquisition and provision in school choice: An experimental study. *Journal of Economic Theory*, *197*, 105345. https://doi.org/10.1016/j.jet.2021.105345

Chen, Y., & He, Y. (2022). Information acquisition and provision in school choice: A theoretical investigation. *Economic Theory*, *74*(1), 293–327. https://doi.org/10.1007/s00199-021-01376-3

Daly, M., Jensen, M. F., & le Maire, D. (2022). University Admission and the Similarity of Fields of Study: Effects on Earnings and Skill Usage. *Labour Economics*, *75*, 102118. https://doi.org/10.1016/j.labeco.2022.102118

DST. (2016). *Overgang fra bachelor til kandidat 2016* (tech. rep.). Statistics Denmark.

Ekbatani, S. (2022). The Cost of Strategic Play in Centralized School Choice Mechanisms. *Working Paper*. https://doi.org/10.2139/ssrn.4134065

Fack, G., Grenet, J., & He, Y. (2019). Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions. *American Economic Review*, *109*(4), 1486–1529.

Gale, D., & Shapley, L. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, *69*(1), 9–15.

Gandil, M. H. (2022). Substitution Effects in College Admissions. *Working Paper*.

Haeringer, G., & Klijn, F. (2009). Constrained school choice [Publisher: Elsevier Inc.]. *Journal of Economic Theory*, *144*(5), 1921–1947. https://doi.org/10.1016/j.jet.2009.05.002

Hassidim, A., Romm, A., & Shorrer, R. I. (2016). "Strategic" Behaviour in a Strategy-Proof Environment. *Working Paper*.

Kapor, A. J., Neilson, C. A., & Zimmerman, S. D. (2020). Heterogeneous Beliefs and School Choice Mechanisms. *American Economic Review*, *110*(5), 1274–1315. https://doi.org/10.1257/aer.20170129

Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of Study, Earnings, and Self-Selection. *The Quarterly Journal of Economics*, *131*(3), 1057–1111. https://doi.org/10.1093/qje/qjw019

Kirkebøen, L. J. (2012). Preferences for lifetime earnings, earnings risk and nonpecuniary attributes in choice of higher education. *Discussion Papers Statistics Norway*, (725).

Larroucau, T., & Rios, I. (2020). Do "Short-List" Students Report Truthfully? Strategic Behavior in the Chilean College Admissions Problem. *Working Paper*.

Larroucau, T., & Rios, I. (2022). Dynamic College Admissions. *Working Paper*.

Manski, C. F. (2004). Measuring Expectations. *Econometrica*, *72*(5), 1329–1376. https://doi.org/10.1111/j.1468-0262.2004.00537.x

Patnaik, A., Wiswall, M., & Zafar, B. (2021). The routledge handbook of the economics of education, ch. College Majors. Routledge. https://doi.org/10.4324/9780429202520-16

Appendices

1.A Linking program identifiers in the application data to DST registers

There is no direct link between the program identifiers in the application data and the program identifiers in the register data, which are a combination of education codes (UDD) and institution codes (INTSNR). I, therefore, have to create one. Before I describe how I generate the link, it is useful to describe a few difficulties I must address to ensure the link is reliable. There are mainly two issues that prevent me from just making a simple merge of applicants from the application data to the register data by personal identification numbers to see which education and institution codes they have in the register data, when they have gotten an offer in the application data. Firstly, some program identifiers in the application data cover multiple program identifiers in the register data and the other way around. Secondly, I can only observe offers in the application data, not which program the applicant enrolls in and the other way around in the register data. To solve the problem, I partly rely on a simple algorithm that, for each program identifier in the application data, finds the share of applicants with an offer who have enrolled in a program in the register data in the current application cycle. I then say a program combination is linked when a sufficiently high share with an offer for a program in the application data is enrolled in a program in the education registry. I further also condition on programs with a sufficiently high number of applicants with an offer.

In practice, I rely solely on the algorithm for programs with a share $x \geq 95\%$ and at least 100 applicants with an offer in the application data. For the remainder of the programs, I define the links by hand, using the share and number of students to guide me.

There is one further difficulty, the education register only contains students who are enrolled in a program by the 1st of October. Therefore it does not contain early dropouts. This should not be a big problem, but it is conceivable that the tendency to drop out early varies across programs, and further, the problem increases in magnitude as the uptake of a program gets smaller. There is, unfortunately, not much I can do to solve this problem, although I try to mitigate it by the heuristic that programs with smaller uptakes should have a higher share enrolled in a given education identifier.

Using this combined approach, I can identify combinations of education and institution codes from the registry data for 582 out of 897 program identifiers in the application data.

1.B List of example programs within fields

Table 1.B.1 contains lists of examples of education programs within the fields I have defined.

Table 1.B.1: Examples of programs in the defined fields

Field	Example programs
Social science	Sociology (KU), Anthropology (AU), Psychology (SDU), Psychology (AAU)
Humanities	English (AU), Danish (SDU), Danish (KU), History (AAU)
Health	Dentistry (KU), Sports Science (KU), Dentistry (AU), Musical Therapy (AAU)
Natural science	Math (AU), Biology (KU), Computer Science (SDU), Physics (KU)
Engineering	Civ. Eng. Energy (AAU), Civ. Eng. Elektronics (DTU), Civ. Eng. Nanotechnology (AAU)
Other Business	Business Law (CBS), International Business (CBS), Marketing and Management Com. (AU), Business Law (AAU)
Education	Pedagogy (KU), Speach Therapy (SDU), Audiology (SDU)
Economics	Economics (KU), Economics (SDU), Economics (AU), and Economics (AAU)
Medicine	Medicine (KU), Medicine (SDU), Medicine (AU), and Medicine (AAU)
Law	Law (KU), Law (AU), Law (AAU), Law (SDU)
Political science	Political science (KU), Political science (SDU), and Political science (AU)
Business	Business Economics (CBS), Business Economics (SDU), Business Economics (AU)

1.C Belief updating

The following algorithm updates applicants' beliefs. The algorithm is based on algorithm 2 in Larroucau and Rios (2022). Larroucau and Rios (2022) use it to update beliefs as students update the information they have about their abilities in under the different policies they evaluate. I have slightly adapted the algorithm to update beliefs given new program capacities.

$\hat{\theta}$ are the estimated preference parameters, ROL^{Full} is the submitted preference ordering for the full sample, $ROL^{Analysis}$ is the preference ordering for the *analysis* sample, C_{old} is the old program capacities, C_{new} is the new program capacities, $score$ is a vector holding applicant scores, S is the number of simulations, μ is the mean of the cutoff distribution, σ is the standard deviation of the cutoff distribution, p is a matrix containing applicants estimated beliefs, and P is a matrix containing the simulated cutoff distribution for all programs.

Algorithm 1 Updating beliefs

Input: $\hat{\theta}$, $\hat{p}^0$, ROL^{Full}, C_{old}, C_{new}, $score$, ϵ_{tol}
Output: p, P
for `each s in S` **do**
  Solve Portfolio Choice problem to get $ROL^{Analysis}$ given ($\hat{\theta}$, $\hat{p}_s^0$)
  Substitute applicants in $ROL^{Analysis}$ into ROL^{Full}
  Bootstrap cutoff distribution to get P^0 given (ROL^{Full},$score$,C_{old})
  Estimate $\hat{\delta^0} \equiv (\hat{\mu}^0, \hat{\sigma}^0)$
end for
Stack $\hat{\delta}^0$ over simulations S
$\delta_{diff} = 2\epsilon_{tol}$, $k = 1$, $\rho = 0.9$
while $\delta_{diff} > \epsilon_{tol}$ **do**
  for `each s in S` **do**
    Solve Portfolio Choice problem to get $ROL^{Analysis}$ given ($\hat{\theta}$, p_s^{k-1})
    Bootstrap cutoff distribution to get $\tilde{P}_s^k$ given (ROL^{Full},$score$,C_{new})
    Estimate updated beliefs $\hat{p}_s^k$
    Take point-wise convex combination of cutoffs $\hat{P}_s^k = \rho^k \hat{P}_s^{k-1} + (1-\rho^k)\tilde{P}_s^k$
    Estimate $\hat{\delta}_s^k \equiv (\hat{\mu}_s^k, \hat{\sigma}_s^k)$
  end for
  Stack $\hat{\delta}^k$ over simulations S

  Compute $\delta_{diff} = \|\hat{\delta}^k - \hat{\delta}^{k-1}\|$, $\hat{p} = \hat{p}^{k-1}$, $k++$
end while

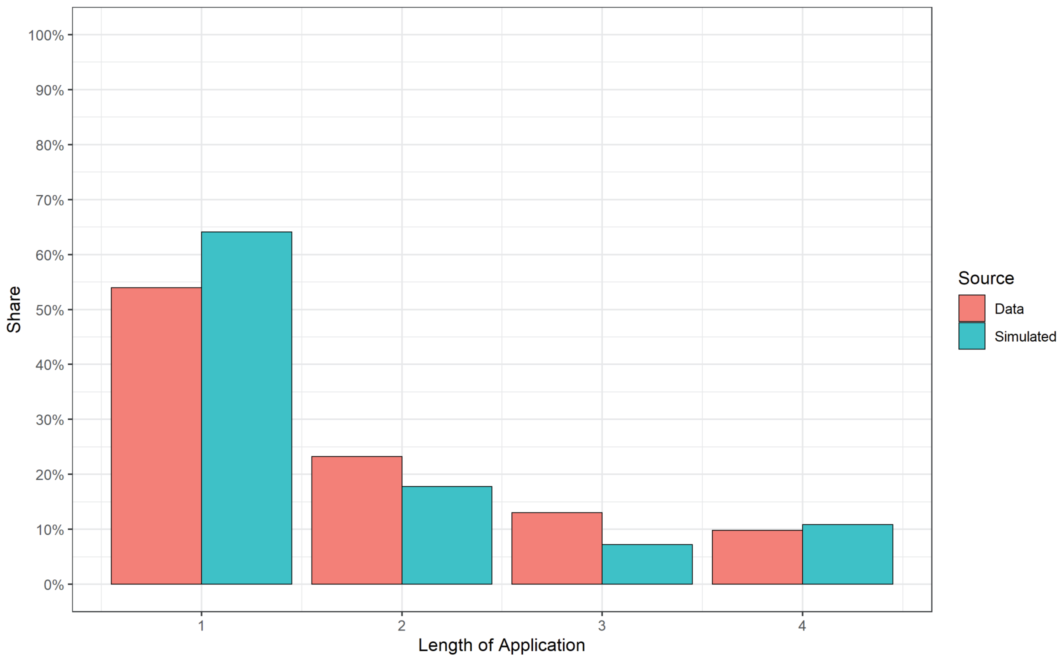
1.D Additional tables and figures

Table 1.D.1: Characteristics of programs in applications by rank

	Rank 4 Mean/Std	Rank 5 Mean/Std	Rank 6 Mean/Std	Rank 7 Mean/Std	Rank 8 Mean/Std
Distance (10 Km)	10.36	9.33	10.16	8.81	10.75
	9.48	8.90	9.07	8.33	8.26
Standardized GPA	-0.36	-0.41	-0.51	-0.38	-0.27
	1.75	1.89	2.02	2.08	1.86
Unemployment (Months)	4.66	4.94	5.13	5.33	5.49
	2.81	2.77	2.81	2.55	3.42
Expected earnings (10,000 DKK)	2.48	2.47	2.45	2.43	2.36
	0.59	0.56	0.56	0.57	0.55
Dispersion of Expected earnings (10,000 DKK)	1.11	1.17	1.18	1.27	1.05
	0.44	0.47	0.49	0.63	0.18
Observations	1,269	494	222	102	53

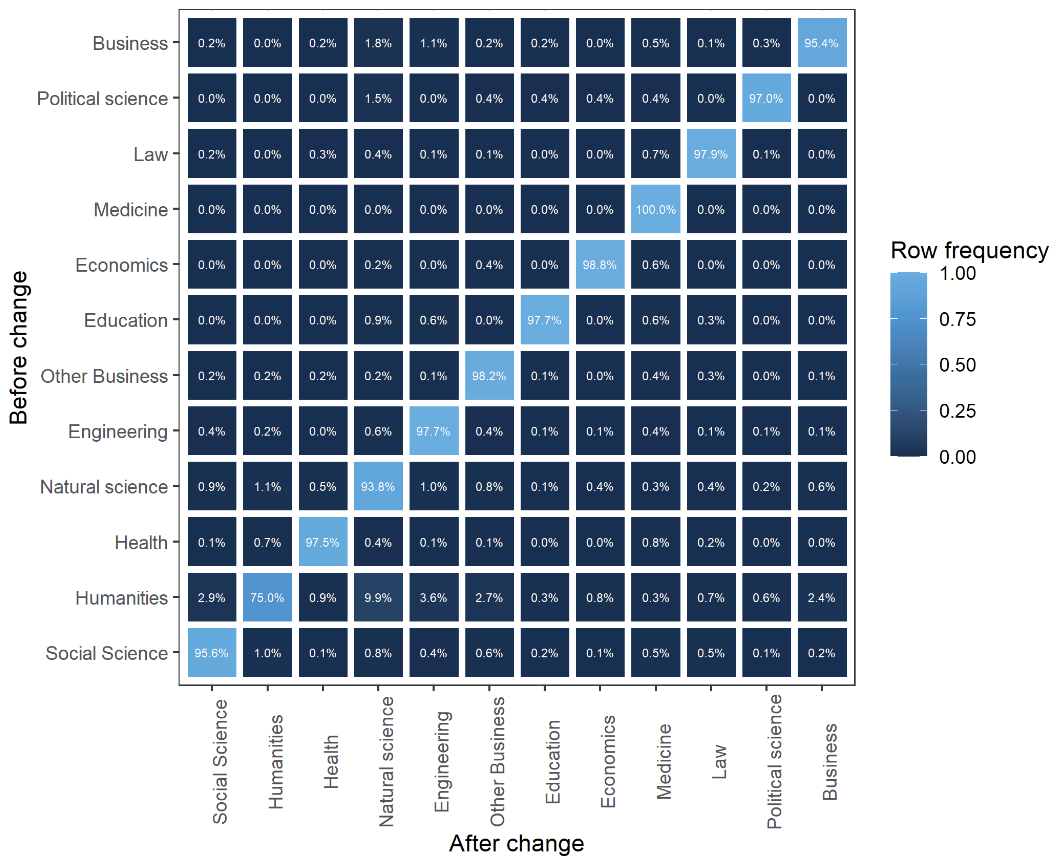
Note: The reported numbers are means (standard deviations in parentheses). The first column reports variable names and units in parentheses. The columns indicate for which rank in the applications the measures are. The number of observations shows the number of applicants with at least the given number of ranks in their application.

Figure 1.D.1: Empirical and simulated share of applicants by the number of programs in their application using beliefs from additional step



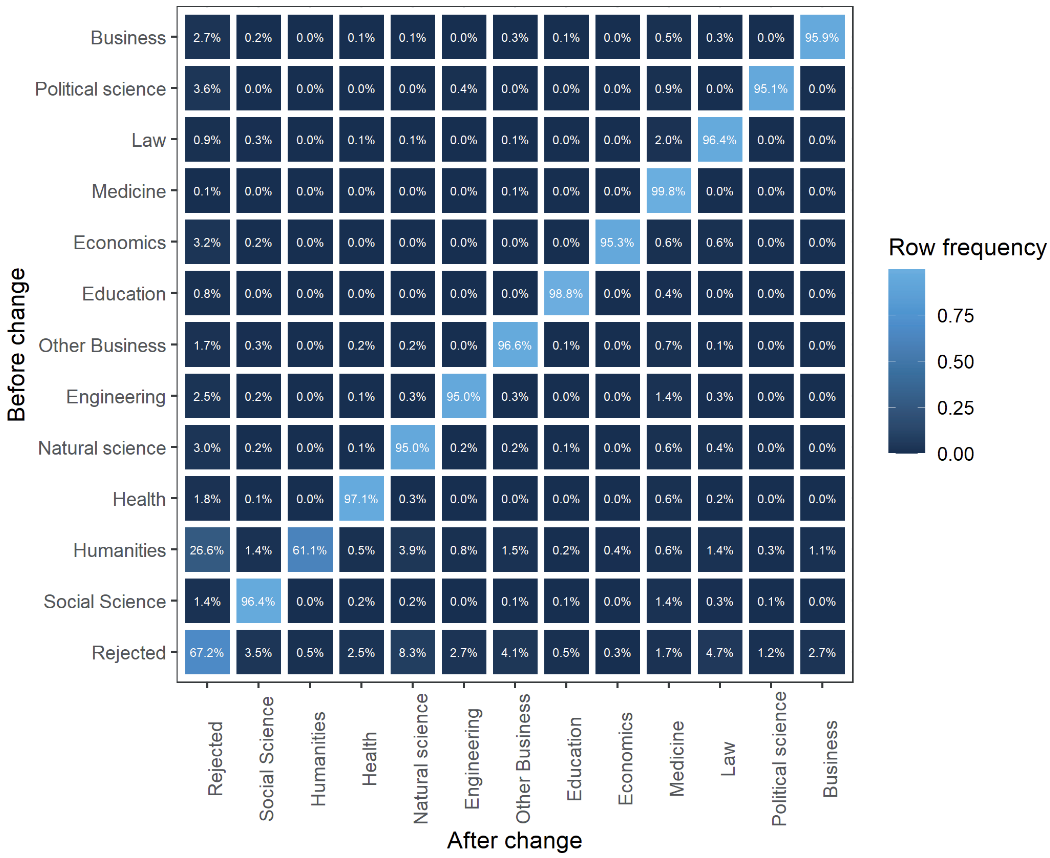
Note: The first axis displays the number of programs in an application (length of the application) and the second axis displays the share of applicants in percentages. The color of the bars indicates the source, red represents the empirical distribution and green indicates the simulated distribution.

Figure 1.D.2: Distribution of top ranked program by field before and after policy change, $\gamma = 0.5$ and applicants update beliefs



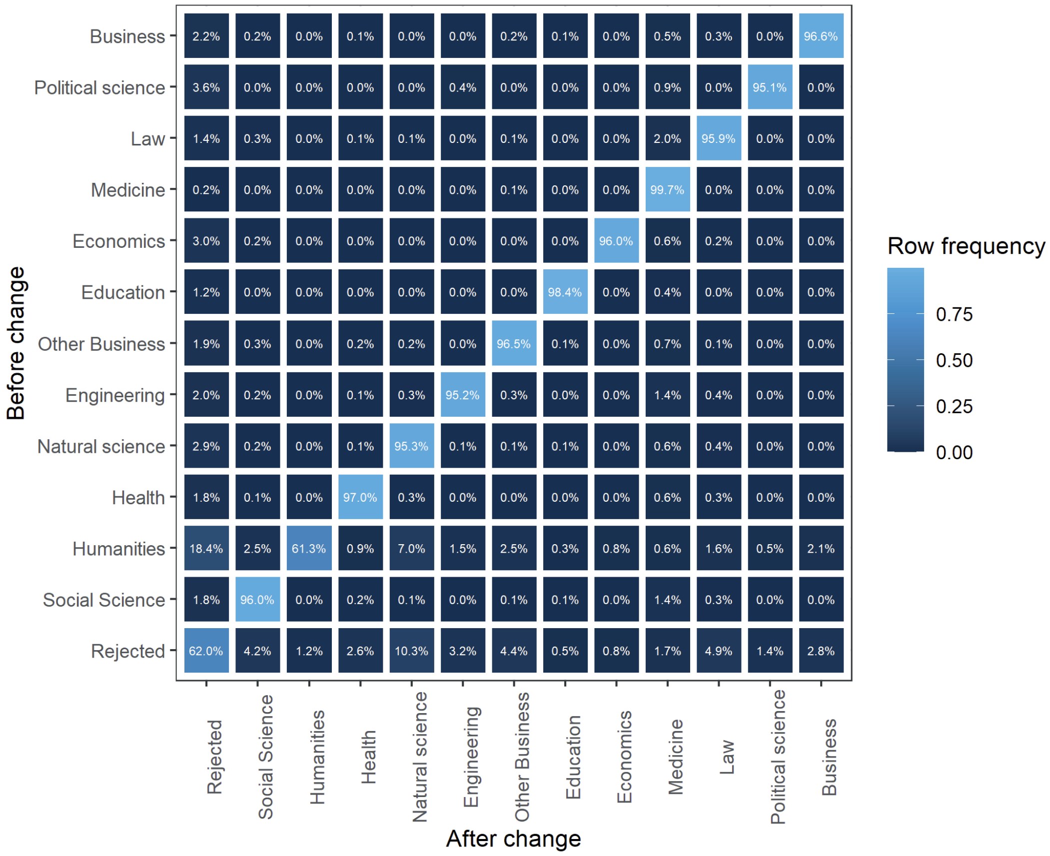
Note: The figure shows the distribution of top ranked programs by the field before and after the policy change. The second axis indicates top ranked field before the change and the first axis indicates top ranked field after the change. The cells are color coded by the share of applications with a given before/after field combination within a before change field, in other words, the percentages on each row sum to 100%. The policy parameter γ gives the fraction of capacities in Humanities which are left, and conversely $1-\gamma$ gives the fraction of capacities in Humanities, which have been redistributed across the other fields.

Figure 1.D.3: Distribution of top ranked program by field before and after policy change, $\gamma = 0.5$ and applicants cannot update beliefs



Note: The figure shows the distribution of top ranked programs by the field before and after the policy change. The second axis indicates top ranked field before the change and the first axis indicates top ranked field after the change. The cells are color coded by the share of applications with a given before/after field combination within a before change field, in other words, the percentages on each row sum to 100%. The policy parameter γ gives the fraction of capacities in Humanities which are left, and conversely $1 - \gamma$ gives the fraction of capacities in Humanities, which have been redistributed across the other fields.

Figure 1.D.4: Distribution of top ranked program by field before and after policy change, $\gamma = 0.5$ and applicants update beliefs



Note: The figure shows the distribution of top ranked programs by the field before and after the policy change. The second axis indicates top ranked field before the change and the first axis indicates top ranked field after the change. The cells are color coded by the share of applications with a given before/after field combination within a before change field, in other words, the percentages on each row sum to 100%. The policy parameter γ gives the fraction of capacities in Humanities which are left, and conversely $1-\gamma$ gives the fraction of capacities in Humanities, which have been redistributed across the other fields.

CHAPTER 2
Absence and Completion among students in Vocational Education

with Fane N. Groes and Edith Madsen

Abstract

We analyze the effect of school absence on program completion among a group of students in Vocational Education and Training (VET) in Denmark. According to human capital theory, being present in class and participating in class activities is an important determinant of human capital formation and therefore the causal effect of absence on educational performance is of interest. To analyze this effect we use data on daily student attendance from the administrative systems of VET schools in combination with register data on completion and student background characteristics. There is a very strong correlation between absence and completion. In order to identify the causal effect of absence on completion, we use local weather conditions such as rain and wind as instruments for student absences. We further introduce a new instrument for absences that uses variation over time in absences for the individual student to support our results. We find that absences during the first two weeks of a 20-week vocational school introductory program has large and significant causal effects on the probability of graduation from the program.

2.1 Introduction

In this paper, we estimate the causal effect of school absence on the probability of program completion for students in vocational education in Denmark. The impact of absence on educational achievement is of obvious interest since most educational programs are based on the assumption that coming to and participating in class is an important input in the skill formation process of the individual student. This together with the fact that student absence in many education programs is substantial underlines the importance of providing evidence of the impact of students being absent from class. Moreover, the existing research on the topic is for students in elementary or academic upper secondary education and in general, there is limited research on student outcomes in vocational education programs.

In our analysis, we consider the basic introductory program in specific vocational programs. This is the focus of our analysis, as there is a very large dropout from this part of the vocational education, see, e.g., Confederation of Danish Employers (2023) and Groes et al. (2021). The basic course is school-based, runs over 20 weeks and it is compulsory for all students. This part of vocational education is taking place at the vocational schools and the classes consists of both practical and theoretical parts. Our outcome is whether the students complete this part of the vocational education. We use data on daily student attendance from the administrative systems at vocational schools in combination with administrative register data from Statistics Denmark on education spells and student background characteristics. Using two different types of instrumental variable estimation strategies, we show that, for students in vocational education, the causal effect of school absence on the probability of program completion is large and significant.

Estimating the causal effect of absences is potentially plagued by endogeneity of absence in that the same unobserved factors that affect individual student absence may also be the ones that affect the student's school outcomes. Examples of such factors are the student's ability level and motivation for learning. Analysis of the causal effect of absence on educational outcomes is challenging especially because of data availability on educational outcomes in combination with student absence and variation in student absence that is uncorrelated with unobserved student factors.

To overcome the potential selection problem that arises from students self-selecting into being absent from school, which might also correlate with completion, we propose two different types of instruments for school absences. Our preferred strategy is a classical instrumental variable, where we use local weather conditions as instruments. The second is a new instrumental variable approach where we use that student absences are measured repeatedly over time giving a panel dataset of this variable.

In our first strategy we use that the number of days with precipitation and high wind leads to more absences among a group of vocational students. Using our weather variables as instruments for absence, we find a causal effect on the compliers of 1.4-1.9, such that a 10-percentage point increase in absence during the first two weeks of school causes between 14-19 percentage point decreases in the probability of completing the 20 weeks introductory program. When analyzing heterogeneous effects, we find that absence among students who do not live with their parents and students with no previous labor market attachment or school enrollment has the highest correlation with our weather instruments. This supports a hypothesis where students with less support for getting out of the door in the morning and students who are not used to having a scheduled start time for the day are the students driving our first stage results.

The second instrument is a panel data instrument, inspired by Arellano and Bover, 1995, where we make use of the fact that even though our outcome measure, completion, is cross-sectional, we have repeated observations over time in absences. The idea is that under some assumptions, we can remove the individual-specific fixed effect from student absences and use this as an instrument for the average individual absence during the first two weeks of school.

Our analysis adds to a small but growing literature estimating the causal effect of school absence on student's educational outcomes. Using panel data which has within student variation over time or between subjects, Cattan et al., 2023 and Liu et al., 2019 find negative effects of absence on short-run academic outcomes and longer-run educational and socioeconomic outcomes for students in elementary, middle, and high school.

Aucejo and Romano, 2016 use variation in flu exposure across counties and time to instrument for absences and find that a reduction in absence increases math and reading scores among students in North Carolina public schools.

We contribute to this literature by analyzing the effect of absences in VET programs on the probability of completion. This is an education area characterized by high dropout and with students coming from low socioeconomic backgrounds. Accordingly this area has a potentially high policy interest as students who do not complete vocational school have a high probability of ending up as unskilled, see Groes et al., 2021.

Closest to our first instrumental variable approach is Goodman, 2014, which considers elementary and secondary school students and instruments absence with snowfall. Goodman, 2014 separates between days with heavy snowfall that causes school closures and days with moderate snowfall that affect individuals' absences. He finds that absences and not school closures affect test scores because school closures affect everyone at the school, and the teachers can change their teaching accordingly. Goodman, 2014 argues that heavy snowfall that causes school closures violates the exclusion restriction as an instrument for student absence because the snowfall affects both absence and the learning process through school closure.[1] In our data, we do not have any days with school closures; thus, our weather instruments affect only individual student absences and not school closures.

Another threat to this identification strategy is if the weather event that causes students to be absent from class also causes them to accumulate less human capital (lower productivity) if coming to class. Several studies have shown that weather affects productivity. Dell et al., 2014 and Park et al., 2020 present evidence that high temperatures affect labor productivity and disrupts learning time and Heissel and Norris, 2018 show that hours of sunlight in the morning before school increases student test scores. Mellon, 2020 surveys the literature using weather as an instrument and encourages a thorough discussion of why the exclusion restriction holds when using weather as an instrument. In our paper, if the weather directly affects student productivity, this would violate our exclusion restriction. We argue that since we use variation in precipitation and wind, which we conjecture only affect how costly it is to get to school, they do not matter for indoor productivity once students are at the school. We believe our results are strengthened by the fact that we use weather over short periods and only throughout Denmark, which does not have large and permanent spatial differences in weather. Furthermore,

[1] Kristensen et al., 2020 and Gottfried, 2009, 2010 use background characteristics to control for selection into absence or instruments absence with distance to school. Both of these approaches possibly suffer from potential bias in estimating the causal effect of absence on school outcomes.

we include school location and month fixed effects in our analysis, such that we use weather variation within the season (students start at different points in time within a given month and across years) and school location. Since Denmark is a country that often experiences precipitation and high wind, all schools are built such that students and teachers can comfortably undertake education during such weather conditions. However, extreme rain may affect how wet students and teachers get during their transport to school, which can last for a while during the day and potentially affect productivity. It could also be the case that heavy rain gives delays in the traffic. In that case it could be that teachers are late to class, which in turn affects teacher productivity. There are no extreme weather occurrences in our sample period and we therefore believe that our exclusion restrictions hold, such that weather does not affect student productivity directly during our sample period.[2] Further, we believe that our exclusion restrictions are more likely to hold compared to studies using health shocks as instruments for absences. This is because learning is less likely to be affected when the weather is bad compared to when experiencing a health shock.

The second instrument we propose is inspired by the panel data literature. We are interested in causally estimating the effect of average absence during the first two weeks of school on the probability of completing an introductory course after 20 weeks. Using the assumption of mean-stationarity from Arellano and Bover, 1995, we assume that we can divide individual weekly absences into an additive individual-specific fixed effect and a time-varying random component that are independent of each other. Under this assumption, we can remove the individual-specific fixed effect by taking the weekly difference, leaving only the time-varying random part of the weekly absences. If we assume no correlation between the de-meaned part of absences and the regression error in the completion regression, we can use it as an instrument for the average absence. Because we only have repeated observations on the explanatory variables, the assumption of mean-stationarity is stronger than the one required in a classical panel data setting with repeated observations on both the explanatory and the dependent variable.

To our knowledge, we are the first to propose an instrument that requires panel data on the explanatory variable but only cross-sectional data on the dependent variable. Besides identifying the causal effect of absence on completion in this paper, we believe this IV approach can be useful in other settings where researchers have repeated observations of the explanatory variable but only one observation of the outcome of interest. Examples could be, measuring the effect of hours spent in daycare on long-run school outcomes (grades, level of education etc.) or the effect of worker absence on the prob-

[2]Sarsons, 2015 shows that there is also threats to identification when using weather as in instrument for the effect of income shocks on conflicts in developing countries. In our setting, we assume that weather does not affect income or any other background characteristics. Auffhammer et al., 2013 survey potential pitfalls when using station level weather data. One concern is that the weather is extrapolated using a grid, which together with the underlying data process of weather causes spatial multi-collinearity that can lead to increased standard errors on the weather. In our data, the weather at the different school locations is indeed highly correlated on a given day. However, we obtain variation in weather across schools for the same starting month because students start at different dates across educations and schools.

ability of job promotion. It is important to emphasize that the method only controls for endogeneity that arises because of an individual-specific fixed effect that is constant over time.

Using the panel data instrument we find that the effect of average absence during the first two weeks on completion is negative such that increases in absences decreases the probability of completion. The panel instrument is currently a work in progress. As we have high share of students with zero weekly absence during the first two weeks, the first-difference across weeks does not remove the fixed effect for these students. We are waiting for better data, which will help us overcome this issue.

To sum up our findings, using precipitation and wind as instruments for absence, we find a large and significant negative effect of absence during the first two weeks on the probability of completion from the second basic course. We support our results with a new panel instrument that is still work in progress. The large effects of absence on program completion from using the weather instruments, suggest high pay-offs for policy interventions encouraging vocational education students to attend classes during the first weeks of school. We leave this for future research.

The paper mainly contributes to three strands of the literature in addition to the bodies of work we have discussed above. First, it contributes to the literature on what causes absence from school. Currie et al., 2009 shows that days with high Carbon Monoxide (CO) increase school absences for public school students in Texas and Zimmer, 2019 finds that visiting a doctor increase the absences among children ages 6 to 13. There is also a small literature on how teacher quality affects student absenteeism where Gershenson, 2016, Tran and Gershenson, 2018, and Liu and Loeb, 2019 show that teachers significantly affect student absences. Our results include all types of absences, both from sickness and absences that do not have a reason attached to them. The first stage estimation using our weather instruments is similar in nature to estimating the effect that Carbon Monoxide has on absences. The absences from doctor visits and teacher quality should not affect the validity of using weather as an instrument as long as neither doctor visits nor teacher quality is related to daily changes in the weather. Our panel instrument uses individual variation over time, so this instrument could potentially use variation from doctor visits, but the fixed effect of teacher quality will be removed when de-meaning student absences. Finally, experimental evidence has found that changing parental beliefs can reduce students' absences in the early grades. Rogers and Feller, 2016 and Robinson et al., 2018 show that a parent-focused intervention on the beliefs about the importance of school attendance and their children's placement in the absentee distribution significantly decreases chronic absenteeism. Due to the large effects, we find of student absences during the first two weeks of enrollment in vocational school, we believe that the vocational education program in Denmark is another obvious place to introduce interventions that reduce student absence.

Second, a complementing strand of the literature analyzes the effect of school days or instruction time on educational outcomes. Analyzing the effect of school days on student performance, Marcotte, 2007, Hansen, 2011, and Marcotte and Hemelt, 2008 use snow days to instrument for school closures and find that number of school days before the exam positively affect the student test scores. Using variation in exam dates and intelligence tests given in the military, Fitzpatrick et al., 2011 and Carlsson et al., 2015 show that more school days before the exam increase students' test scores. Groppo and Kraehnert, 2017 use difference-in-differences to show that severe winters in Mongolia affect the medium- and long-run education outcomes, while Craig and Martin, 2019 find that eliminating student suspension increases student test scores. A different way of increasing students' time in school is by expanding the instruction time during the day. Using reforms that change the school day length, Dominguez. Patricio and Ruffini, 2018 show there is a positive effect on educational attainment by increasing the school day in Chilean elementary and secondary schools and Lavy, 2020 shows that increased time at school with more school resources lead to increase student achievement. Finally, Lavy, 2015, Rivkin and Schiman, 2015, and Bingley et al., 2018 utilize within student differences in taught hours across subjects to find a positive effect on test scores and that these positive effects vary by the classroom environment. Our identification strategy is also closely related to this last strand of literature. However, instead of using within student variation and weather as exogenous changes in school length, we use the variation to predict student absence, holding the school length and instruction time constant. Goodman, 2014 and Aucejo and Romano, 2016 both show that the effect of reducing absence is larger than a comparable effect of increasing instruction time. With this in mind, the fact that we find particularly large effects of absences suggests we would not find as large effects if we instead increased the instruction time in the vocational education program.

Finally, this paper also contributes to the literature on student dropout from vocational schools. Denmark, like Germany, Switzerland, and Austria, has what Eichhorst et al., 2015 refers to as a dual system for the VET, which is characterized by a high degree of formalization, vocational schools that provide the school-based part of the dual apprenticeship, and accreditation of the firms that students train in during their apprenticeship. Vocational schools in Denmark have a large dropout rate and are constantly under pressure to increase student completion.

At the same time, the students at vocational schools have relatively high absences at an average rate of 7.5 percent of days during the first two weeks of school increasing to around 12 percent at the end of week three. By providing causal evidence of student absence on the probability of completion, our result hopefully contributes to a better understanding of how to help students in vocational education.

By analyzing the causal effect of absence on the completion probability, we contribute to the understanding of the vocational education system, where the economic literature mainly has concentrated on the effect of obtaining a vocational education on labor market outcomes (see Hanushek, 2012, Hanushek et al., 2017, Hampf and Woessmann, 2017, Bertrand et al., 2021, and Silliman and Virtanen, 2022). Stratton et al., 2017 analyze the probability of completion from vocational school, taking selection on grades from mandatory school into account. They find that prior math scores are particularly important for completion, which we also find in our results and therefore include as one of the background characteristics in our estimation.

The remainder of the paper is organized as follows: Section 2.2 describes the Danish VET system. Section 2.3 describes the data we use along with our sample selection criteria. Section 2.4 shows descriptive statistics. Section 2.5 describes our empirical strategy. Section 2.6 describes our results. Finally, section 2.8 concludes.

2.2 Institutional setting

In Denmark, the present vocational education and training system (VET) was introduced in 2014 and implemented in August 2015. The VET system consists of more than 100 types of vocational programs; a few examples of specific educations are carpenter, electrician, and hairdresser.

The Danish VET consists of a basic program and the main program. The basic program takes one year and is divided into two separate courses. Each takes 20 weeks (excluding holidays) and consists of theoretical and practical classes where all teaching is exclusively at the vocational school, which is in contrast to the main program where the majority of the time is spent on internships away from the school. The main program will typically take 3-3.5 years and alternate between school courses and apprenticeships, where the latter takes place in a specific company or organization. There is an exam at the end of both basic courses and the end of the main program.

The first basic course (GF1)[3] is very broad and mainly meant to help the students choose a specific education (for example, carpenter), which they will follow from the second basic course and on through the main program. In general, class attendance is compulsory for VET students, and the teachers and student counselors have to monitor student absences and follow up on students being absent from classes. However, there are no strict rules regarding the maximum absence allowed for a VET student.

This project is concerned with students enrolled in the second basic course (GF2). Two possible channels allow students to enter the second basic course. The first channel is through the first basic course, reserved for students who attend VET within two years after completing the mandatory 9th or voluntary 10th grade. The second channel is for students who completed their lower secondary education more than two years earlier. These students start directly on the second part of the basic course, i.e., they skip the first basic course. This means they choose a specific education at the beginning instead of after half a year. Students over 25 years old attend the special VET for adults, which can be with or without the second basic course and possibly with a shorter main program.[4] For older students, the vocational school will decide whether the person needs the second basic course or whether he/she can start directly in the main program. The special VET for adults takes up a large share of the students in the VET system. Having many older students enrolled is very different from the upper secondary education (high school) that prepares students for higher education programs. Here, most students are under 20 years old, as most recently completed lower secondary education.

[3] We use GF1 and first basic course interchangeably for the remainder of the paper.

[4] The composition of 25+ study depends highly on their prior education and work experience. An example of this could be a person who has worked several years as an unskilled carpenter and wants to have formal education as a carpenter.

The different types of students and their different pathways through the VET system imply that many vocational schools have two starting dates for the second basic course within a year: January and August. For example, students who recently completed lower secondary school will typically start the first basic course in August and therefore start the second basic course in January the year after. The other students begin immediately with the second basic course and, therefore, can begin their education in either January or August. Altogether, this means that the mix of students in the second basic course can differ substantially depending on the time of the year.

Another feature of the Danish VET system is that it offers a VET education combining general upper secondary education and vocational education and training (EUX). EUX qualifies students for jobs as skilled workers and gives direct access to higher education within a wide range of programs. In our analysis, we exclude students who combine upper secondary education with VET.[5]

2.3 Data Sources and Sample Selection

For our analysis, we use three different data sources. We have collected data on absence from eight vocational schools that we merge with the Danish register data to get student completion and background characteristics. We also merge our data with daily meteorological observations and a measure of distance from the school.

2.3.1 VET School data

We collected our primary data on absence at eight primarily large vocational schools in Denmark that offer technical education for this project.[6]

The data covers all individual spells at the schools for 2015-2019, including education, institution, and start date.[7]

The data also contains information on the daily student class schedule, the number of scheduled hours, how many hours the student attended, and the reason for not attending to some extent. Schools classify student absence as either ”excused” (for example illness) or ”not excused” (reason unknown). In this project, we do not distinguish between the types of absence but instead consider total absence. We do this primarily because many of the student absences have missing information on the cause, but also because the type of absence is likely to be subject to miss-classification error. For example, a

[5]A full overview of the different pathways through the Danish VET system is found in appendix figure 2.A.1

[6]The schools were selected if they had a carpenter education to ensure we had one large education represented at all the schools, to ensure that the number of observations was high.

[7]The data also includes the end date and reason for the end of a spell. We choose not to use this information from the school data because the end of a spell date is unreliable since schools sometimes overwrite the data if a student starts a new education. Further, the variable which contains the reason for ending a spell has many missing observations.

student might report being ill while the reason for absence is something else. In addition, our starting point is that absence in itself, no matter the cause disrupts the instruction and training of the student.

We consider absence within the first two weeks of classes in the second part of the basic course. We do this for two reasons. Firstly, to avoid sample size issues. Students drop out continuously, so we will not have information on some students for the last weeks. Secondly, it is to avoid having to deal with shocks to the teaching/learning process, which could correlate with both the dropout decision and absence. We elaborate further on the second reason in section 2.5. Using the collected data, we construct our measures of interest: *individual percentage of absence during the first two weeks of the second basic course.*

2.3.2 Register data

We combine the school data with the register data from Statistics Denmark. We match the school data by individual id, education, institution, and starting date to the Danish Student Register (KOTRE), where we observe, by dates and school, each educational enrolment spell and any credentials obtained from the students' educations. We can match 95.6 percent of the student spells from the school data to the register data. For the majority of students, we match their spell by the exact matriculation date. For the remaining students, we create matches by allowing the matriculation dates in the school data to differ by up to 21 days compared with the register data, provided the education and institution are the same and a sufficient number of observations in our data share this matriculation date (we have chosen 15). For the analysis, we use the exact first day as the start date, which is the information we have from the school data.

Completion information is, as mentioned earlier, only complete for some spells in the school data, so we use the information in KOTRE to define our outcome, namely completing the second basic course within seven months. In addition, we use the enrolment status and the matriculation date from the register data to define the timing of the completion.

For background characteristics of the students in the second basic course, we combine data from five different administrative registers. First, from the Danish Student Register, we observe, by dates, each educational spell the students have ever enrolled in and any credentials obtained from these. We use this data to construct the highest completed degree at the time of the first enrollment in the second basic course and an indicator for students coming from the first basic course as well as previous unsuccessful attempts at the second basic course. For individuals who completed 9th grade after 2002, we observe the grades used to qualify for vocational and high school. Among all the grades from the 9th grade, we chose the Danish and math grade received at the end-of-year exam, which are national exams given to all 9th-grade students. Second, from the demographics register (BEF), we extract background characteristics on age, sex, and immigration status. Third, we use the Integrated Database for Labor Market Research (IDA) to observe parents' labor market status. Fourth, we obtain information on the parents' highest completed education level from the education register (UDDA). Lastly, we obtain information on students' primary employment status during the year before enrolment in the second basic course from the AKM register.

2.3.3 Weather and distance data

We combine our data with weather data from the Danish Meteorological Institute's (DMI) online service. DMI collects daily information on, e.g., wind, precipitation, and temperatures from sensors dispersed over Denmark, so we have access to information on the weather conditions at the municipality of the schools.[8]

For our analysis, we use aggregated measures of our weather observations. However, since we observe both the weather and the students' classes daily, we can use the students' schedules during the first two weeks of school and, for each student, merge the weather data by date and municipality for the days that the students have scheduled classes. Since we can match our weather measures to the school days by date, students with different start dates also have different weather observations during their first two weeks of classes, even if they go to the same school. Further, students who, e.g., only have nine days of scheduled classes are only matched to the weather during those nine days.

We create two primary weather measures for our analysis. We use daily millimeters (mm) of precipitation and high wind speeds, which we average individually over the first two weeks of scheduled school days. The chosen measures are mm of daily precipitation and the highest average wind speeds in meters per second (ms) in a 10 min. interval.

Lastly, we combine our data with the distances between the parish in which the student lives and the parish the school is located in. We define distance as the bird's flight distance from one parish's centroid to another parish's centroid. We include the distance as a measure of the distance the student has to travel to school, and since most students live relatively close to their respective schools, some of the usual concerns with bird's flight distance, such as crossing the sea, are not a problem. It is worth noting that since we use parish-level measures, we lose the within-parish-level variation in distance, although this is not a big concern as parishes are relatively small areas, and there is still much between parishes variation.

2.3.4 Sample Selection

For our analysis, we restrict our sample to the period after a reform to the VET area in Denmark, which was implemented on August 1, 2015. The reform introduced a new structure, where the first part of the VET program is now divided into a first and a second basic course. Furthermore, to follow the students up to seven months after their

[8]To generate municipality-level information, DMI uses information from the weather stations, which are dispersed across the country. Based on the measures obtained from the weather stations, DMI first generates a fine weather grid for the whole country and aggregates it to the municipality level. Thejill et al., 2020 gives a more detailed description of the methods used to generate the grid and municipality aggregation.

enrollment in the second basic course, we select the last month of enrolment as August 2018 and focus on the two regular starts of the education during August and January.[9] Finally, since our collected sample of schools primarily includes technical schools, we restrict our sample to only include the seven largest VET programs *carpenter, joiner, electrician, smith, plumber, house painter,* and *mason*.

The VET students in the second basic course have a high dropout rate, and the schools have students dropping out every week. Therefore, we select students that stay enrolled in the course during the first six weeks and have at least 15 scheduled hours during each of the first two weeks to avoid including students who have already dropped out in our measure of student absence. This leaves us with a final analysis sample of 5782 observations or 85,915 daily x student observations.

2.4 Descriptives

Table 1 shows means and standard deviations for the students' characteristics and our measures of interest, which are absence and completion. Column (1) presents the full sample and Columns (2) and (3) present the sample split by whether or not the students attended the first basic course prior to their enrollment in the second basic course. The chosen VET educations are very male-dominated with on average 89 percent of students being male. Most of the students in our analysis sample are natives (90 percent). The average 9th-grade math and Danish grades are 4.4 and 4.7, which are lower than the grades in the population, where the average overall grade is around seven (7.3 in 2018 (The Danish Ministry of Education, 2018)). Further, around half of students (59%) live with their parents, although this varies a lot by whether the students have attended the first basic course or not. Students have, on average, ten days of school during the first two weeks, corresponding to full school weeks, excluding weekends. Further, on average, students have 5 hours of classes, excluding breaks, each day of which they are absent for around half an hour. During the first two weeks, 43 percent of students have zero hours of absence. Seventy percent of the students in our sample completed within seven months, and this pattern is similar for students with and without the first basic course. Thirty-two percent of the students in our analysis sample have attended the first basic course.

Comparing the students with a prior first basic course in Column (2) to students who start directly at the second basic course in Column (3), we see that they are more likely to be males (96 compared to 86 percent) and on average younger (17 years compared to 22 years). Students with a prior first basic course are also more likely to be natives (94 percent) compared to students without (89 percent), where the difference mainly is in the share of 1st generation immigrants (1 compared to 7 percent) while the share of second-generation immigrants is similar.

Beyond the individual student characteristics, we also include parents' education and employment status, presented in Table 2. Most noticeably, 46 percent of students have

[9]This is an end of sample restriction made at the time we started the project.

a parent where Vocational training is their highest education. In addition, a little over 70 percent of students have mothers and fathers who are employed, around 7 percent have mothers or fathers registered as unemployed, 17 percent have mothers or fathers who are out of the labor force, and 4-5 percent have mothers or fathers with missing observations.

2.4.1 Pattern in Absences

We next turn to describe the patterns of absence in our data. Table 3 shows the weekly patterns of absences for the first three weeks of school. In panel A, we see that the percentage of absent students increases over the three weeks, with 27, 41, and 48 percent absent for at least one hour during weeks 1, 2, and 3. The average hours of absence also increase over the first three weeks, both unconditionally and conditionally on having some absence, while the scheduled number of hours per week remains the same. When we analyze the effect of absence on completion, we use absence in percent of scheduled hours, which is presented in Panel B. Also here, we see that the students increase their absence in percent over the first three weeks, both conditionally and unconditionally on having some absence during a given week, although the increase from week 2 to 3, conditionally on having positive hours, is smaller than from week 1 to 2. In panel C, we see how the individuals' absences correlate over weeks. The first row in Panel C presents the weekly probability of having some absence for students who had some absence in week 1. For these students, the probability of absence during week 1 is 100 percent by definition. For weeks 2 and 3, they are 43 percent and 38 percent. Comparing these percentages to the overall sample in row 1 of panel A, we can see that when we condition on students with some absence in week 1, the share of students with some absence in week 2 is similar to the unconditional share, while the share of students with some absence in week 3 is lower than the unconditional share. This pattern is different for students with some absence during weeks 2 and 3, presented in rows 2 and 3 in Panel C. For both these types of students, having some absence in either week two or week three is associated with a higher share of students with some absence than the unconditional share in the other two weeks. These statistics suggest that the absences during week 1 are more random than the absences during week two and week three, which seems to be more individual-specific.

We use average absence during the first two weeks as our variable of interest when we analyze the effect of absence on completion. We saw from Table 1 that 43 percent of students have zero absences during the first two weeks. In order to investigate the heterogeneity in absence across students, we consider the distribution of average absence within the first two weeks of class. In Figure 1 we show the distribution of average individual absences during the first two weeks in bins of 10 percentage point absences. Figure 1 shows some heterogeneity in student absence. Approximately 65 percent of students are absent for 0 to 10 percent of the hours during the first two weeks, 18 percent of students are absent for 10 to 20 percent of the hours, 8 percent of students are absent for between 20 to 30 percent of the hours, and few students are absent for than

Table 1: Descriptive student characteristics

	Full Mean (SD)	First Basic Course Mean (SD)	No First Basic Course Mean (SD)
Male	0.890	0.959	0.858
	(0.312)	(0.198)	(0.349)
Age	20.635	16.967	22.337
	(5.261)	(0.740)	(5.580)
log(distance (km))	2.408	2.505	2.363
	(1.104)	(0.909)	(1.181)
Native	0.901	0.942	0.883
	(0.298)	(0.233)	(0.322)
Immigrant	0.050	0.013	0.068
	(0.219)	(0.111)	(0.252)
2nd gen. immigrant	0.048	0.045	0.050
	(0.214)	(0.208)	(0.217)
Lives with parents (< 25)	0.590	0.955	0.421
	(0.492)	(0.207)	(0.494)
Unemployed	0.016	0.000	0.024
	(0.126)	(0.000)	(0.152)
In school	0.617	1.000	0.440
	(0.486)	(0.000)	(0.496)
Other	0.123	0.000	0.180
	(0.328)	(0.000)	(0.384)
Employed	0.203	0.000	0.298
	(0.402)	(0.000)	(0.457)
Benefits recipient	0.040	0.000	0.059
	(0.197)	(0.000)	(0.236)
9th grade danish score	4.358	4.245	4.419
	(2.724)	(2.145)	(2.990)
9th grade math score	4.760	4.701	4.793
	(3.081)	(2.603)	(3.314)
Daily hours (2 weeks)	5.189	5.198	5.184
	(0.465)	(0.466)	(0.464)
Daily hours absence (2 weeks)	0.390	0.354	0.406
	(0.652)	(0.570)	(0.686)
Days school (2 weeks)	9.924	9.972	9.901
	(0.338)	(0.208)	(0.381)
Absence>0 (2 weeks)	0.527	0.529	0.526
	(0.499)	(0.499)	(0.499)
GF1	0.317	1.000	0.000
	(0.465)	(0.000)	(0.000)
Graduated within 7 months	0.695	0.712	0.687
	(0.460)	(0.453)	(0.464)
Observations	5782	1833	3949

Note: The tabel reports mean characteristics and standard deviations in parentheses for the full sample and the sample split by whether the students have a prior spell in the first basic course (second column), or not (third column). We do not observe 9th grade Danish and Math grades for all students, the reasons for this are that the grades were not recorded in the register before 2002 and also some students have not attended 9th grade in Denmark.

Table 2: Descriptive parental characteristics

	Full Mean (SD)	First Basic Course Mean (SD)	No First Basic Course Mean (SD)
Highest completed parental education			
Missing	0.053	0.017	0.069
	(0.223)	(0.131)	(0.253)
Primary school	0.249	0.259	0.244
	(0.432)	(0.438)	(0.430)
Highschool	0.036	0.027	0.040
	(0.185)	(0.163)	(0.195)
Vocational school	0.465	0.556	0.422
	(0.499)	(0.497)	(0.494)
Short track HE	0.051	0.056	0.049
	(0.220)	(0.229)	(0.216)
Prof. BA	0.097	0.063	0.112
	(0.295)	(0.244)	(0.315)
BA	0.006	0.003	0.007
	(0.075)	(0.057)	(0.082)
MA/PhD	0.045	0.017	0.058
	(0.208)	(0.131)	(0.234)
Fathers employment status			
Employed	0.729	0.806	0.693
	(0.444)	(0.395)	(0.461)
Unemployed	0.065	0.063	0.065
	(0.246)	(0.244)	(0.247)
Out of the labor force	0.171	0.124	0.192
	(0.376)	(0.330)	(0.394)
Missing	0.035	0.006	0.049
	(0.185)	(0.077)	(0.216)
Mothers employment status			
Employed	0.707	0.785	0.671
	(0.455)	(0.411)	(0.470)
Unemployed	0.075	0.065	0.080
	(0.264)	(0.247)	(0.271)
Out of the labor force	0.162	0.136	0.175
	(0.369)	(0.343)	(0.380)
Missing	0.055	0.014	0.074
	(0.229)	(0.118)	(0.263)
Observations	5782	1833	3949

Note: The tabel reports mean characteristics and standard deviations in parentheses for the full sample and the sample split by whether the students have a prior spell in the first basic course (second column), or not (third column).

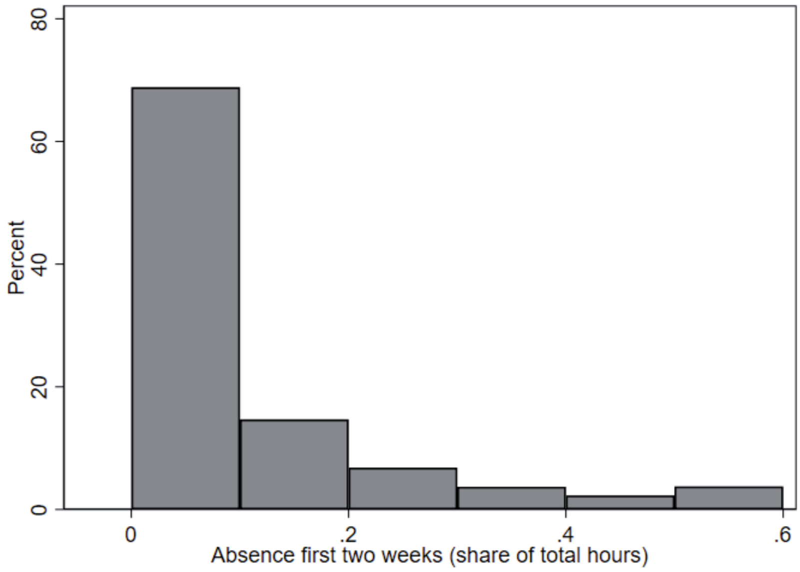
Table 3: Descriptive patterns in student absence

Panel A	Week 1 Mean(SD)	Week 2 Mean(SD)	Week 3 Mean(SD)
Hours absence> 0	0.268	0.413	0.484
	(0.443)	(0.492)	(0.500)
Hours absence	1.335	2.531	2.979
	(3.268)	(4.650)	(4.783)
Hours absence (cond. > 0)	4.983	6.130	6.156
	(4.658)	(5.505)	(5.264)
Hours	25.705	25.790	25.741
	(2.603)	(2.842)	(3.033)
Panel B	Week 1 Mean (SD)	Week 2 Mean (SD)	Week 3 Mean (SD)
Pct. absence	0.052	0.098	0.116
	(0.127)	(0.181)	(0.187)
Conditional pct. absence	0.194	0.238	0.240
	(0.181)	(0.214)	(0.206)
Panel C: Conditoinal on hours absence> 0	Share Week 1	Share Week 2	Share Week 3
Week 1	1.000	0.428	0.388
Week 2	0.659	1.000	0.588
Week 3	0.702	0.689	1.000
Observations with positive absence	1549	2387	2798

Note: Panel A reports means and standard deviations in parentheses for different measures of absence and hours for the first three weeks of the second basic course. Panel B reports means and standard deviations in parentheses for weekly pct. absence and weekly pct. absence conditional on having positive absence. Panel C reports the conditional fractions of observations with positive weekly absence split by the three first weeks. The rows indicate which week is conditioned on and the columns indicate which week the fraction is for. Panel C also reports the number of observations with positive absence in a given week. The table is based on the analysis sample containing 5782 observations.

30 percent the hours. Figures 2 and 3 show the distribution of hours by education and by whether the student had prior enrollment in the first basic course. Both Figures show the same patterns as Figure 1, such that there does not seem to be much difference in the absence distribution within the first two weeks across education or prior enrollment in GF1.

Figure 1: Distribution of hours absent as a share of total hours for the first two weeks

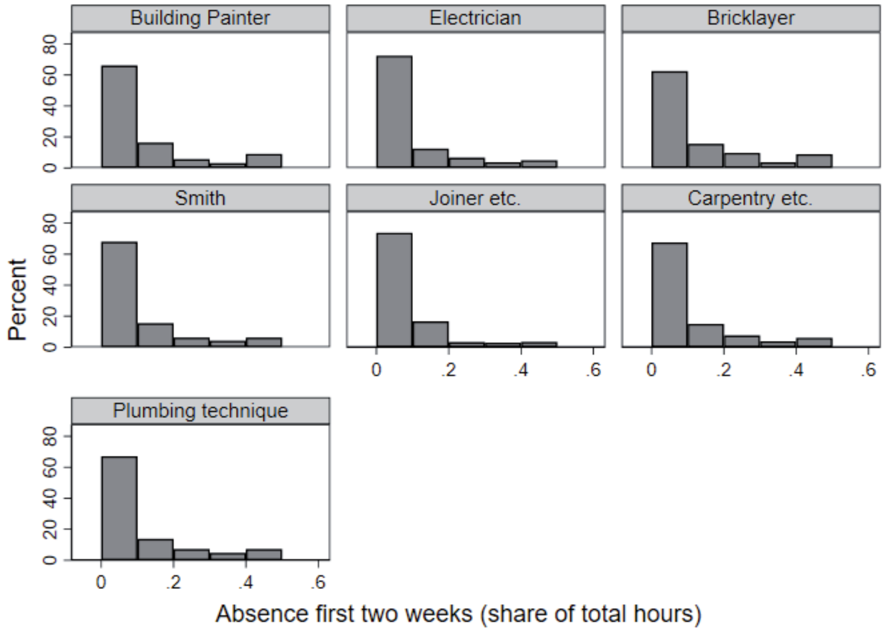


Note: The figure has the share of students on the second axis and the percent hours absent on the first axis. The first bare only includes students with zero absence.

We use the individual weekly variation in absences for the panel data instrument. Table 3 describes the average weekly variation, and Figure 4 shows the daily absences for all students and students with some absence during the day. In order to avoid selection issues, the sample consists of students that have scheduled class every day Monday through Friday during the first four weeks (20 weekdays). Figure 4 shows that the unconditional daily absence increases over the first 20 days, although at a decreasing rate. On the first day of school, around 4 percent of students are absent, and on day 20, this has increased to around 17 percent. The percent of daily hours absent for students with some absence during the day stays somewhat constant at around 65 percent of daily absence, meaning the average student who is absent is so for around 4 out of 6 hours during the day.

The patterns in daily absence are much the same across all educations as illustrated in Figure 5. However, when we look at the absence by whether the student has a prior enrollment in the first basic course or not, Figures 1.6(a) and 1.6(b) show that the older students starting directly on the second basic course grow to have a higher percent daily absences and more absence on the days where they have some absence. So on days the older students are absent, they tend to be absent for more hours than the younger

Figure 2: Distribution of hours absent as a share of total hours for the first two weeks by education

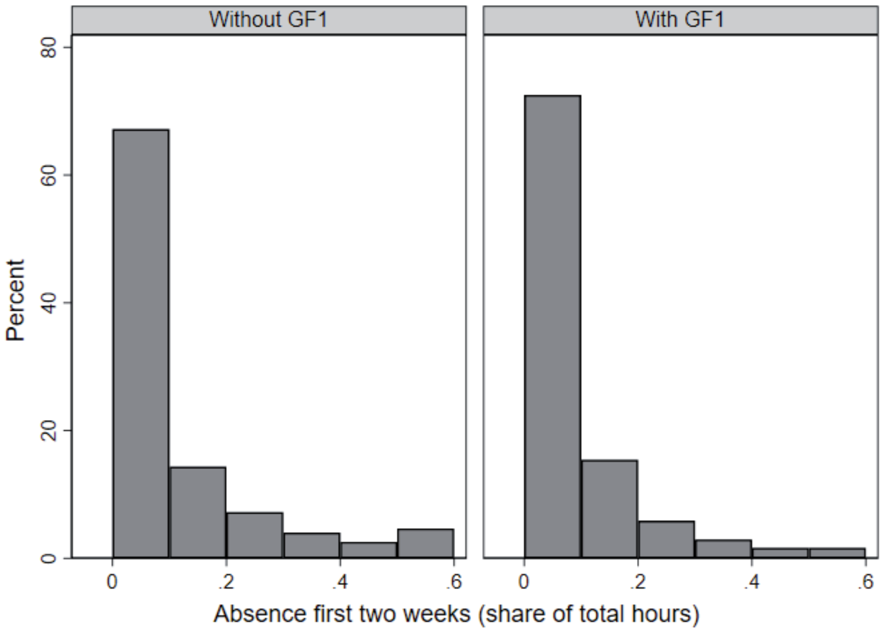


Note: The figure consists of 7 panels, one for each of the educations in our analysis sample. The panels all have the share of students on the second axis and the percent of hours absent on the first axis. The first bar in each panel only contains students with zero absence.

students illustrating that just using an indicator for daily absence may not capture this type of heterogeneity in hours of daily absence. The last curiosity is that the percent of daily absence conditional on having some absence varies over the days. The daily absences are higher around days 1, 5, 10, 15, and 20. To the extent that students start on a Monday, this shows that students who have some absence during the day have more hours absent on Mondays and Fridays.

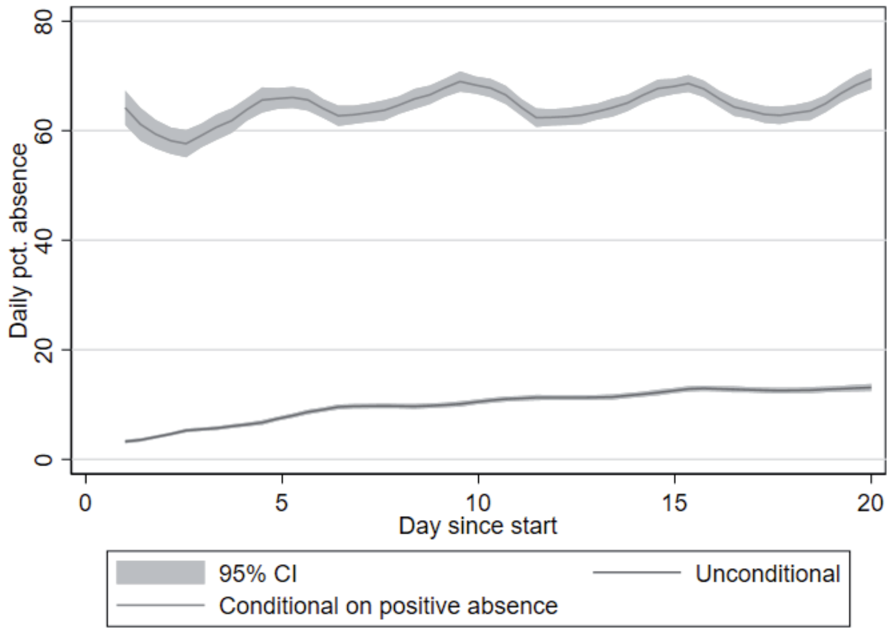
Finally, Figure 7 shows the relationship between the average absence during the first two weeks of school and the completion of the second basic course. We see that the relationship is very strong. Approximately 75 percent of the students with no or a very low level of absence complete the education, whereas the number is approximately 55 percent for those with an absence level of 20 percent (corresponding to two full days of absence during the first two weeks if the hours are equally distributed across weekdays). Part of this relationship can most likely be explained by the fact that students with high ability, effort, motivation, etc., are more likely to perform well in their courses and, in the end, complete the second basic course and are more likely to attend class. Therefore, we expect only part of the relationship to be a causal effect of absence on completion.

Figure 3: Distribution of hours absent as a share of total hours for the first two weeks by whether the students come from the first basic course or not



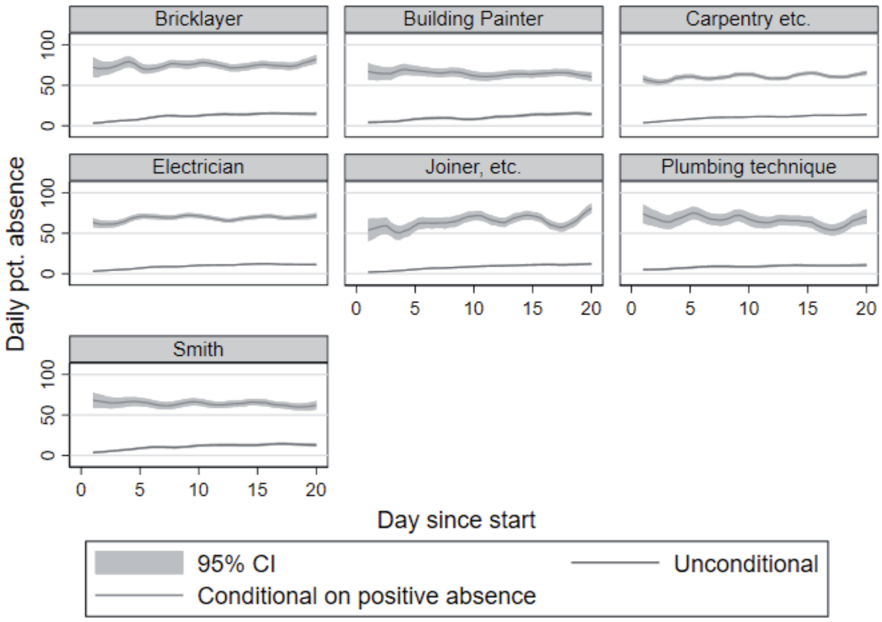
Note: The figure consists of two panels, the first is for students who start directly on the second basic course and the second is for students from the first basic course. The panels both have the share of students on the second axis and the percent of hours absent on the first axis. The first bar in each panel only contains students with zero absence.

Figure 4: Daily percent absence

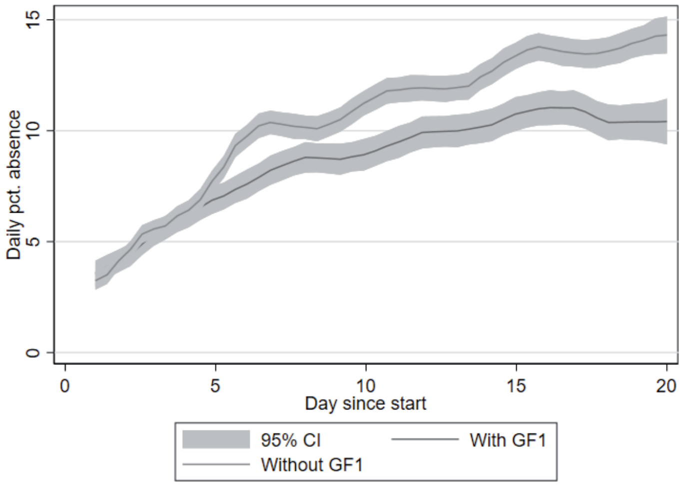


Note: The figure has the daily percent absence on the second axis and days since the matriculation date on the first axis.

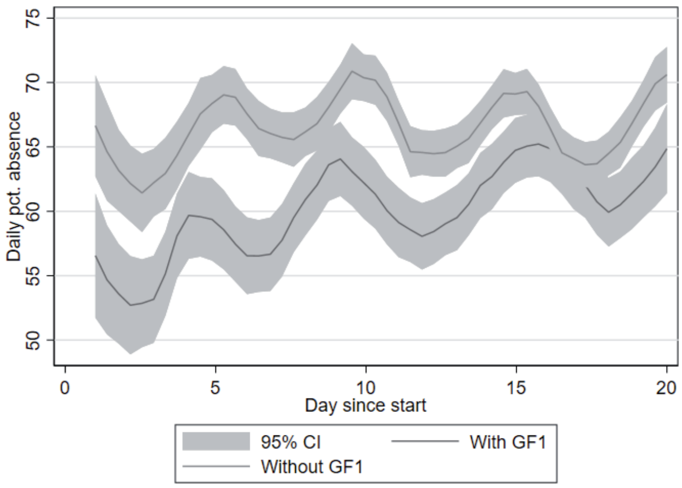
Figure 5: Daily percent absence by education



Note: The figure consists of 7 panels, one for each of the educations in our analysis sample. The panels have daily percent absence on the second axis and days since the matriculation date on the first axis.



(a) Unconditional

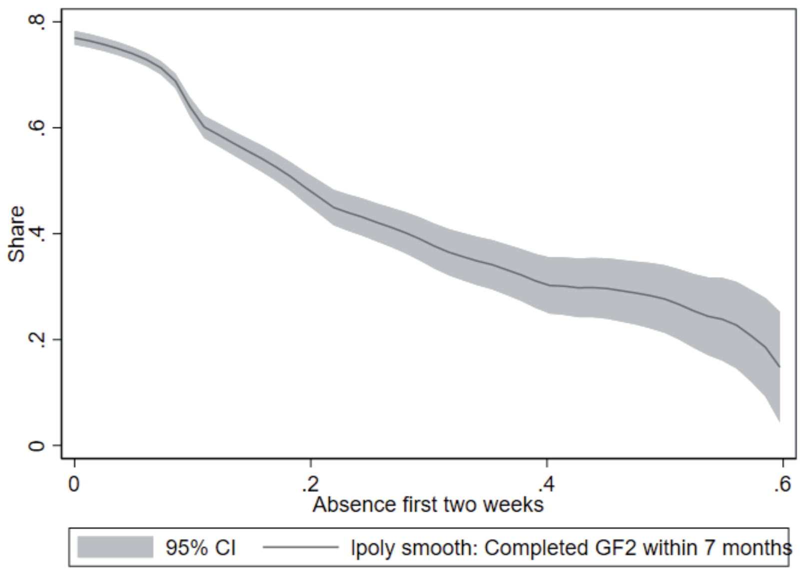


(b) Conditional on positive daily absence

Figure 6: Daily percentage of absence by prior GF1 or not

Note: The figure consists of two panels. Both panels have days since initial enrollment on the first axis and daily percent of absence on the second axis. Panel (a) displays the daily unconditional percent of absence for students with a prior first basic course (red line) and without a prior first basic course (blue line). Panel (b) displays the daily conditional percent absence for students with a prior first basic course (dark grey line) and without a prior first basic course (light gray line).

Figure 7: Relationship between total absence during the first two weeks and completion approximated by polynomial regression



Note: The figure has the share of students completed within seven months on the second axis and the percent of absence during the first two weeks on the first axis. The line is from the kernel weighted polynomial regression of absence on completion.

2.5 Empirical strategy

In this section, we describe our empirical strategy. According to human capital theory, being present in class and participating in class activities is an important determinant of human capital formation, and therefore, the causal effect of absence on educational performance is of interest. In the previous section, we saw a strong correlation between being absent from class within the first two weeks and the likelihood of completing the second basic course. We expect part of the correlation between absence and course completion to be confounded by underlying unobserved individual-specific factors such as student ability and motivation, as it is likely that high-ability and high-motivation students are more likely to come to class and to complete the course. Therefore, in order to identify the causal effect of absence from class on the likelihood of completing the second basic course, we consider two identification strategies:

a) Exogenous instruments based on meteorological measurements of weather conditions

b) A panel data instrument for absence

We explain the strategies in detail below.

2.5.1 IV methodology using weather

Our main identification strategy relies on instruments based on daily meteorological measurements of weather conditions at the vocational school locations. We use measures of precipitation and wind in the first two weeks of the course to match the period for which we measure absence. These precipitation and wind measures vary with year and month (January and August weeks) and with the location of the school as well as with the exact starting dates of the course (some start Monday in a given week, others start Wednesday, and some start the week after). In addition, we control for year, month, location, and education group fixed effects. Controlling for the combination of fixed effects rules out the case where weather variations affecting all students' absence in only one year or month (or location or education) are driving the results.

To identify the effects of absence on graduation, we make use of a standard IV setup, where we will use different specifications of weather during the first two weeks of the second basic course as instrumental variables; hence we will run a 2SLS regression model to estimate the effect of absence during the first two weeks of the second basic course on the probability of graduating with the following first- and second-stage equations

$$\bar{a}_i = \delta_0 + Z_i\delta_1 + X_i\delta_2 + \lambda_m + \lambda_y + \lambda_s + \lambda_e + v_i \tag{2.1}$$

$$y_i = \beta_0 + \beta_1\bar{a}_i + X_i\beta_2 + \gamma_m + \gamma_y + \gamma_s + \gamma_e + \epsilon_i, \tag{2.2}$$

where y_i is completion of the second basic course for individual i at the end of the course. The explanatory variable of interest is $\bar{a}_i$, which is the individual-level average absence within the first two weeks of the course. We are mainly interested in the parameter β_1, which is the marginal effect of absence on completion. X_i is a vector of exogenous covariates such as age, gender, immigrant status, parental education, and parental labor market attachment Finally, γ_m, γ_y, γ_s, and γ_e are respectively month, year, school, and education fixed effects.

Equation 2.2 is our equation of interest. We use this equation for the OLS regressions and the second stage with instrumented average absence. For the instrument, we use exogenous variation in the average weather, Z_i, during the first two weeks of school to predict average absence. More specifically, using weather during the first two weeks of school, we use the percentage of days with precipitation above 3 mm as our first instrument and the percentage of days with wind above 11 m/s for a 10 minute interval as our second instrument. We use the instruments separately and together, as well as interacted. We show that our result is robust to different definitions (cut-off values) of the weather variables. This is explained in more detail in section 2.6.2.

2.5.2 Panel data instrument

In addition to the main identification strategy in Section 2.5.1, we supplement our analysis with instruments based on the availability of repeated observations over time of individual absences (on a daily or weekly basis). The idea is that under specific assumptions, we can remove the individual-specific fixed effects from absence, and then under the assumption that there is no remaining correlation with the error term in the regression equation, this provides an instrument for a more aggregated measure of absence, e.g., the difference in weekly absence as an instrument for biweekly average absence. Altogether, the underlying assumptions are stronger than those required in the case with panel data observations of both the dependent and explanatory variables. This type of assumption is well-known from the dynamic panel data literature, where it is used to estimate the autoregressive parameter in a dynamic panel data model with individual-specific fixed effects. In that setting, the estimation is done by using lagged first-differences as instruments for the equation in levels. The assumption is known as mean-stationarity and introduced in Arellano and Bover, 1995.

We consider the following regression model:

$$y_{iT} = \beta_0 + \beta_1 \bar{a}_i + X_i \beta_2 + \eta_i + u_{iT} \tag{2.3}$$

where y_i is completion of the second basic course for individual i at the end of the course after T time periods (20 weeks of school). Note that equation 2.3 is similar to equation 2.2 except for notation where we have omitted the fixed effects in the notation (they are included in the vector of exogenous covariates, X_i) and the error term is now split in two parts.

The regression error consists of the two terms η_i and u_{iT}. Note that we cannot distinguish the two error terms from each other in the cross-section setting. We assume that average absence $\bar{a}_i$ is independent of u_{iT} but can be correlated with η_i. The first error term η_i captures individual-specific fixed effects such as student ability and the part of student motivation that is constant over time. The term η_i affects both absence and the completion of the second basic course. The second error term u_{iT} captures all other unobserved parts of completion that, by assumption, are independent of average individual absence within the first two weeks. The term u_{iT} can contain both individual-specific time-constant effects and individual time-varying shocks that happen to the learning process during the 20 weeks of the course (for example, teacher quality and things going on in and outside the class). Part of the identifying assumption is that there is no reverse effect from learning process shocks to absence, i.e., $\bar{a}_i$ is independent of u_{iT}, and this is also the reason for considering absence during the first two weeks of class. We conjecture that after some time in class, we are more likely to find reverse causality in the sense that shocks to the teaching/learning process that impact the gains from the teaching in the class also has an impact on the likelihood of coming to class in later weeks. We find it plausible that this reverse effect is not present within the first two weeks of class, where students are finding out what the course is about and attendance is possibly less influenced by realizations of gains from attending class. Another part of the identifying assumption is that the effect of absence in weeks one and two is the same such that there is only one endogenous variable in the regression equation of interest.

Letting a_{it} be absence within a week, t, we assume

$$a_{it} = \widetilde{a}_{it} + \alpha_i \tag{2.4}$$

where $\widetilde{a}_{it}$ and α_i are independent of each other with mean zero and $\widetilde{a}_{it}$ is independent of the regression errors η_i and u_{iT}. The individual-specific effects α_i and η_i can depend on each other, and the dependency between absence and the regression error happens through this term. Altogether this implies that the first-differences $\Delta a_{i2} = a_{i2} - a_{i1} = \tilde{a}_{i2} - \tilde{a}_{i1}$ are independent of η_i such that $E\left[\Delta a_{i2}(\eta_i + u_{iT})\right] = 0$. In addition we have that $E\left[(a_{i1} + a_{i2})(a_{i2} - a_{i1})\right] = E(\widetilde{a}_{i2}^2) - E(\widetilde{a}_{i1}^2)$ such that Δa_{i2} is a valid instrument for $(a_{i1} + a_{i2})$ if the second order moments of $\widetilde{a}_{it}$ are not constant over time. We can test whether this holds in the first-stage regression. In addition, any function of Δa_{i2} can be used as instrument for $(a_{i1} + a_{i2})$, for example Δa_{i2}^2 (this requires assumptions on third order moments) or $|\Delta a_{i2}|$. Note that $\tilde{a}_{i1}$ and $\tilde{a}_{i2}$ can be dependent over time. Altogether the assumption on absence a_{it} means that a_{it} must be additive in the individual-specific effect α_i such that we can remove the term by subtracting the individual-specific mean. As mentioned above, this assumption is referred to as mean-stationarity in the panel data literature, see Arellano and Bover, 1995. The assumption rules out that low- and high-ability students can have different trends in absence over time, and therefore it cannot be the case that low-ability students increase absence over time more than high-ability students but they can have permanent high level of absence. In the situation with more than two observations over time of absence that are exogenous except for the α_i-part, the instruments will be the individual-mean corrected absences (or first differences of absences) at the different periods. The approach requires that absence is a continuous variable such that it can be expressed in a linear additive form as in equation 2.4. We discuss this in more details in the next section.

2.6 Results

This section presents our results regarding the effect of absence during the first two weeks of school on the probability of completing the second basic course. We first present the baseline OLS estimates of equation 2.2. Second, we present the results from using our chosen instruments based on meteorological observations, presented in section 2.5.1, and include first-stage estimates as well. In all the estimations, we use the average percent of hours absent during the first two weeks of school as our variable of interest and an indicator of whether the student has completed the second basic course within 7 months after graduation as our outcome variable. We also show the results are robust to using percent days with some absence rather than percent hours absent during the first two weeks.

Third, we support our results from the main specifications with the panel data instrument from section 2.5.2. We see that the results are qualitatively similar to our main results, using meteorological weather observations as instruments. We include the first-stage estimates and robustness checks where we condition the sample on students with some absence.

2.6.1 OLS results

Table 4 shows the OLS results from estimating the effect of absence within the first two weeks on completion from equation 2.2. The table presents results for three different groups where panel A includes the full sample, panel B only includes students *with* a prior spell in the first basic course and panel C only includes students *without* a prior spell in the first basic course. The table's columns show how the coefficient on absence changes as we add controls and fixed effects sequentially. The relationship between absence during the first two weeks and completion probability is highly significant and stable around a coefficient of approx -1 for all three groups. In column (7), we see that including all the controls and month, year, education group, and school fixed effects, the OLS estimates for all students indicate that a 10 percentage point increase in hours absent (corresponding to one full day of absence) during the first two weeks is associated with a 9.1 percentage point lower probability of completing within seven months.

In appendix table 2.A.1 we present the raw OLS results for percent days with some absence instead of percent hours of absence. The estimates are similar in magnitude and size as the ones for percent hours of absence in table 4 and they remain stable as we include more control variables.

The censoring of absence at zero is a concern for our panel data instrument because it introduces a non-linearity in absences. Therefore, we also present all results conditional on students with some absence in at least two weeks, leaving us with approximately 18 percent of the sample. Table 5 presents the OLS estimates for absence within the first two weeks as explanatory variable when we condition on only including students with some absence in both weeks. In table 5, we see that the association between completion

Table 4: The Effect of pct. Hours Absence during the first two weeks on Graduation: OLS results

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)	Graduated (5)	Graduated (6)	Graduated (7)
Panel A: All							
Pct. absence	-1.039***	-0.916***	-0.906***	-1.037***	-1.032***	-1.030***	-0.909***
	(0.046)	(0.047)	(0.048)	(0.047)	(0.048)	(0.049)	(0.047)
Observations	5782	5782	5782	5782	5782	5782	5782
R-squared	0.081	0.144	0.154	0.081	0.095	0.096	0.174
Panel B: With GF1							
Pct. absence	-1.030***	-0.902***	-0.908***	-1.033***	-1.061***	-1.065***	-0.948***
	(0.132)	(0.119)	(0.119)	(0.132)	(0.134)	(0.135)	(0.119)
Observations	1833	1833	1833	1833	1833	1833	1833
R-squared	0.062	0.133	0.146	0.063	0.087	0.088	0.183
Panel C: Without GF1							
Pct. absence	-1.039***	-0.909***	-0.898***	-1.036***	-1.001***	-0.997***	-0.883***
	(0.055)	(0.061)	(0.061)	(0.056)	(0.055)	(0.056)	(0.061)
Observations	3949	3949	3949	3949	3949	3949	3949
R-squared	0.088	0.159	0.169	0.091	0.104	0.107	0.189
Student controls	No	Yes	Yes	No	No	No	Yes
Parent controls	No	No	Yes	No	No	No	Yes
Year/Month FE	No	No	No	Yes	No	Yes	Yes
School/Education FE	No	No	No	No	Yes	Yes	Yes

Notes: The table shows the raw OLS estimates for pct. absence during the first two weeks on graduation (0/1). The bottom panel indicates if student controls (age, immigration status, log distance, male, 9th grade danish score, 9th grade math score, and labour market attachment prior to enrolment), parent controls (highest completed parental education, mothers and fathers labour market attachment), year and month fixed effects, and school and education fixed effects are included in the model in the respective column. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no prior GF1. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

and absence within the first two weeks is lower when we condition on some absence than the unconditional absence in table 4. As we saw in table 4, the point estimate in table 5 is only slightly smaller (towards zero) when we include control variables.

The fact that all the estimates are robust to the inclusion of the different control variables suggests that the students only, to a small extent, select into being absent based on observed characteristics. However, the OLS effect of absence on completion might still suffer from potential bias due to unobserved omitted factors, in particular unobserved individual heterogeneity that is constant over time, which is why we continue with our strategy for the panel data instrument.

2.6.2 Weather Instrument Results

2.6.2.1 First-stage Results, Weather Instrument

Table 6 shows the first-stage estimates for the different weather instruments in column (1) percent days with over 3 mm precipitation during the first two weeks, in column (2) percent days with average wind speeds above 11 ms for 10 minutes, in column (3) the two instruments jointly, and in column (4) the two instruments jointly along with their interaction. Panel A reports the first-stage estimates for the full analysis sample. We

Table 5: The Effect of pct. Hours Absence during the first two weeks on Graduation, conditional on two weeks positive absence: OLS results

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)	Graduated (5)	Graduated (6)	Graduated (7)
Panel A: All							
Pct. absence	-0.642***	-0.635***	-0.605***	-0.629***	-0.624***	-0.614***	-0.587***
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Observations	1021	1021	1021	1021	1021	1021	1021
R-squared	0.050	0.160	0.192	0.055	0.069	0.073	0.224
Panel B: With GF1							
Pct. absence	-0.761***	-0.829***	-0.800***	-0.763***	-0.812***	-0.832***	-0.873***
	(0.19)	(0.18)	(0.19)	(0.19)	(0.19)	(0.20)	(0.19)
Observations	341	341	341	341	341	341	341
R-squared	0.047	0.141	0.166	0.051	0.081	0.087	0.234
Panel C: without GF1							
Pct. absence	-0.577***	-0.571***	-0.547***	-0.578***	-0.530***	-0.531***	-0.515***
	(0.09)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.08)
Observations	680	680	680	680	680	680	680
R-squared	0.045	0.189	0.229	0.053	0.073	0.078	0.268
Student controls	No	Yes	Yes	No	No	No	Yes
Parent controls	No	No	Yes	No	No	No	Yes
Year/Month FE	No	No	No	Yes	No	Yes	Yes
School/Education FE	No	No	No	No	Yes	Yes	Yes

Notes: The table shows the raw OLS estimates for pct. absence during the first two weeks, conditional on some absence in both weeks, on graduation (0/1). The bottom panel indicates if student controls (age, immigration status, log distance, male, 9th grade danish score, 9th grade math score, and labour market attachment prior to enrolment), parent controls (highest completed parental education, mothers and fathers labour market attachment), year and month fixed effects, and school and education fixed effects are included in the models in the respective column. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no proir GF1. Robust standard errors are in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

see that wind is separately significant on the 5% significance level, while precipitation separately and their interaction is significant on the 10% level for the full sample. We see the same pattern, although with larger and more significant coefficients and F-values, in panel C, which contains the estimates for the students who did not attend the first basic course before their enrollment in the second basic course. In panel B, which reports estimates for the students who attended the first basic course before their enrollment in the second basic course, we see insignificant coefficients for all columns. Further, all the columns have very low F-values indicating that our weather measures are not suitable candidates as instruments for the students who attended the first basic course prior to their enrollment in the second basic course. The fact that none of the coefficients are significant leads us to conclude that the weather instruments are only appropriate for the students without a prior enrollment in the first basic course, and we, therefore, primarily focus on this group in our main analyses.[10] The first-stage results for panel C all point

[10] A potential reason why the older students react more to weather conditions is that they have been out of school for a number of years and many of them have also not held a job prior to enrolling in vocational school. Because they are less used to getting out the door in the morning, the marginal student may be more likely to be affected by the weather relative to the younger students who have not tried anything but attending school every day. This may also be true for the students who live without their parents relative to students not yet having moved away from their parents. We will analyze these heterogeneous effects in Section 2.6.2.4.

in the expected direction, namely that students are more absent over the first two weeks when there are more days with precipitation or high wind speeds.

Table 6: First-stage estimates for the different specifications of weather instruments, percent hours absent used as outcome (2 weeks)

	Pct. hours absent (1)	Pct. hours absent (2)	Pct. hours absent (3)	Pct. hours absent (4)
Panel A: All				
Pct. days with precipitation > 3mm	0.032		0.022	0.004
	(0.020)		(0.022)	(0.019)
Pct. days with wind>11ms		0.039**	0.034**	0.029*
		(0.015)	(0.017)	(0.017)
Precipitation X wind				0.076
				(0.057)
Observations	5782	5782	5782	5782
F	2.42	6.51	3.82	3.06
Panel B: With GF1				
Pct. days with precipitation > 3mm	-0.035		-0.033	-0.033
	(0.029)		(0.028)	(0.037)
Pct. days with wind>11ms		0.023	0.018	0.018
		(0.028)	(0.026)	(0.036)
Precipitation X wind				0.000
				(0.092)
Observations	1833	1833	1833	1833
F	1.49	0.70	0.88	0.68
Panel C: Without GF1				
Pct. days with precipitation>3mm	0.082***		0.058**	0.019
	(0.026)		(0.027)	(0.023)
Pct. days with wind > 11ms		0.09***	0.073***	0.063***
		(0.020)	(0.021)	(0.021)
Precipitation X wind				0.172*
				(0.090)
Observations	3949	3949	3949	3949
r2				
F	9.76	19.51	12.59	11.27

Notes: The table shows the first-stage estimates for the different specifications of the weather instruments on our endogenous measure of interest, pct. hourss absent during the first two weeks. All models contain additional controls (age, immigration status, log distance, and male), year and month fixed effects, and school and education fixed effects. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no prior GF1. F-tests for joint significance are reported for each panel. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

To show that these first stage estimates are not only occurring for our particular values of precipitation and wind, tables 2.A.3 and 2.A.5 respectively show the first-stage estimates for a range of precipitation and wind specifications for the students without a prior GF1 spell. Table 2.A.3 shows that for any cutoff of precipitation, the first stage estimates are significant at the 1 percent level however, the F-values for the first stage are higher for smaller values of the precipitation cutoff. The F-value is highest when we count days with above 1 mm of precipitation. For days with wind speed above a certain cutoff, table 2.A.5 also shows significant first stage estimates for a range of different wind speeds and with F-values from 10 to 19 for wind speeds from 10 ms to 13 ms.

To further explore the variation in precipitation and wind, figures 2.A.2 and 2.A.3 show the share of students exposed to the different cutoff values. Figure 2.A.2 shows that, respectively, 65 and 80 percent of students without and with GF1 experience at least one day with precipitation above 1 mm. For our chosen value of 3 mm these numbers are 70 and 50 percent of students without and with GF1. The share of students experiencing high winds are somewhat higher with around 97 and 98 percent of students without and with GF1 experiencing at least one day with wind above 11 ms.

2.6.2.2 Main Results: Weather Instruments

Table 7 contains our main results, the IV estimates of the effect of being absent during the first two weeks on the probability of graduating based on our second-stage specification in equation 2.2 with the weather instruments. As mentioned in section 2.6.2.1, we only report results for students without a prior spell in the first basic course, as the first-stage check showed that our weather instruments are not suitable for the group with a prior spell in the first basic course. Therefore, we display the second-stage estimates for the joint analysis sample and students with a prior spell in the first basic course in appendix table 2.A.2.

Column (1) of table 7 reports the second-stage estimate for precipitation as an instrument for percent hours absent. We see that students are estimated to be 1.95 percentage points less likely to graduate when they are 1 percentage point more absent in hours; This roughly corresponds to a 20 percentage point decrease in the probability of graduating when they have one additional day of absence during the first two weeks of courses.[11] The pattern for columns (2)-(4) are similar, although the estimated coefficients are slightly smaller in absolute size, with $1.06-1.36$ percentage points. The coefficient in column (2), with our measure of wind as an instrument, is not significant, although when coupled or interacted with precipitation in columns (3) and (4) the coefficient on absence is significant. Further, the first-stage F-values range from $9.76-19.51$, and we can reject the null on the 1% significance level for the Anderson Rubin wald test for column (1) and at the 5% significance level for column (3) and (4) .

We can again explore how the estimates vary over the precipitation and wind cutoff values. In tables 2.A.4 and 2.A.6 we show the second-stage estimates for different precipitation and wind specifications and in table 2.A.7 we show a matrix of second-stage estimates for the different joint precipitation and wind specifications. Table 2.A.4 show that the second-stage estimates are significant for counting days with 1, 3, 4 and 5 mm precipitation and that our chosen specification of 3 mm has the highest significance level at 5 percent. The coefficients that are significant at the 10 percent level vary 1.69 to 2.19, such that our chosen specification with a 2SLS estimate of 1.95 is somewhere in the middle. For different specifications of wind, table 2.A.6 shows that other cutoffs than 11 ms, indeed have significant second-stage estimates, such that future research should perhaps focus on using a cutoff of 12 or 13 ms rather than the 11 that we chose in the main specification. Table 2.A.7 shows the ranges of second-stage estimates when both precipitation and wind are included. Also here we see that we can reject the null at a lower significance level for the Anderson Rubin wald test if we use a cutoff of 12 ms or

[11] To roughly translate the interpretation of the coefficient from a 1 percentage point increase in absence to one more day of absence, we can multiply the estimated coefficients with 10, which table 2.4 shows is the average number of school days during the first two weeks.

Table 7: Effect of absence on probability of graduation: 2SLS with weather as IV for absence (2 weeks)

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)
Pct. hours absent	-1.952**	-1.059	-1.404**	-1.357**
	(0.928)	(0.817)	(0.636)	(0.611)
Observations	3949	3949	3949	3949
First-stage F	9.76	19.51	12.59	11.27
AR p-value	0.004	0.205	0.018	0.045
Precipitation $> 3mm$	Yes	No	Yes	Yes
Wind $> 11ms$	No	Yes	Yes	Yes
Precipitation×Wind	No	No	No	Yes

The table reports second-stage estimates. The dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

13 ms for 3 mm of precipitation. If we use days with 13 ms wind and days with 3 mm precipitation, the 2SLS coefficient is 1.69 and is significant at the 1 percent level with an F-value of 13.26.

2.6.2.3 Robustness Results: Weather Instrument

Our results are robust to the definition of absence. We run the same estimations with percent days with some absence rather than percent of hours absent. The idea being that days with precipitation and wind will affect either being late for school or being absent the entire day. Categorizing percent days with any absence during the first two weeks of school is a variable from zero to 100, where being absent half an hour or the full day, will give the students 10 percentage point (1 day out of ten) more absence. This is the same way we define the instruments. Table 2.A.8 shows the first-stage estimates for students without GF1 when we use percent days with positive absence. We see that both days with precipitation and days with wind are significant at the 5 percent level although with lower F-statistics than when we use percent hours absence. Using both instruments does not give any additional significance. Together the results show that days with precipitation and wind better predict variation in percent hours of absence rather than days with any absence. This could be explained, for example, if students have a tendency to miss the first class on all days, independent of the weather but that bad weather makes the students miss the entire day. This would give a higher percentage absence on the days with bad weather and would therefore make the weather variable have higher correlation with percent hours of absence rather than days with some absence. Using absence in percent days, table 2.A.9 shows the second-stage estimates of students without a prior GF1 enrollment. The estimates are significant at the 10 percent level and have slightly higher point estimates than when we use percent hours absent. This is such that 10 percentage point more absence (e.g. 1 day more absent during the first 10 days of school) translate into 20 percentage point lower probability of completing the GF2 course.

2.6.2.4 Heterogeneous effects: Weather Instrument

The fact that the weather instrument only affects percent hours absent for the students without a prior GF1 enrollment can be due to many different reasons. As we saw in section 2.4 the two groups with and without a prior GF1 are different on many observable characteristics. Two of these characteristics are the fact that only 42.1% percent of the students without a GF1 do not live with their parents (compared to 95.5% of students with a GF1) and that one quarter have neither worked nor been enrolled in school the year prior to enrollment. In this section, we show how the results vary across these two characteristics for the students without a GF1. The idea being that we want to check if weather affects absence more for students who do not live with their parents (who can encourage them to go to school even if it rains) and students who are not used to having a time to start the day because they have not been in school or at work the year prior to enrollment. Besides differing between students with and without GF1, our hypothesis is that these two characteristics are also likely to correlate with students who are more likely to be affected by weather conditions when they decide to go to school in the morning. We will test this below. Finally, we show heterogeneous results across math grades from 9th grade to test if weather is a stronger instrument for students with low or high grades.

The results on the heterogeneous effects are presented in tables 2.A.10, 2.A.11, and 2.A.12. The first row in each table presents the second stage estimates, the second row the OLS estimates and row three and four the first stage estimates for the two instruments. When we divide students by living with or without parents, in table 2.A.10, we see that the OLS estimates do not differ much across the groups. However, for the precipitation instrument, the results are driven by students not living with their parents. For this group the first-stage coefficients are much larger than for the group of students living with their parents. Further, the F-statistic is 24 for the precipitation instrument, which is also much larger than for the group of students living with their parents who only have an F-statistic of 1. Together, the results from table 2.A.10 show that our hypothesis holds for students not living with their parents as weather is a stronger predictor for their absence, which affects completion.

For students with and without a prior stable attachment (prior enrollment in school or employed), shown in table 2.A.11, the OLS estimates are larger for the group of students without a prior stable attachment indicating that the correlation between absence and completion is stronger for the group of students with no stable attachment. The first stage estimates show that weather predicts absence for both groups but the correlation is stronger for the group of students with no stable attachment. In this sense our hypothesis seems to also hold for students without a prior attachment, although there is no significant effect on completion for students without a prior attachment, most likely due to the smaller sample size. The first-stage estimates translate into significant second-stage estimates for the group with a stable prior stable attachment, although with higher point estimates. We consider this as work in progress, and will look further into this in the future.

Finally, when dividing the sample by high and low math grades from 9th grade, table 2.A.12 shows that the OLS estimates are almost the same for high and low grade students. In the first-stage we see that the weather is more correlated with absence for the low grade students, but the second stage estimates are insignificant. For the high grade students, the second-stage estimates are significant and around twice as large in magnitude as our main results, but the F-statistics of these regressions are low and we therefore do not want to interpret further on these estimates.

2.6.3 Panel Data Instrument Results

2.6.3.1 First-stage Results, Panel Data Instrument

To support our main analysis using different weather measures as an instrument, we implement the same analysis using our panel data instruments.

The panel data instrument is the first difference in average absence between two weeks and functions of the first difference that are the difference squared, the first difference and the difference squared together, and the absolute difference.

Table 8 shows the first-stage estimates from the four different specifications of our instrument for average absence during the first two weeks. The table reports the results for the full sample in panel A, students with a prior first basic course in panel B, and students without a prior first basic course in panel C. All of the instrument specifications correlate significantly with average absence, and if we focus on panel A, the reported F-tests are all above 100. However, a significant driver of the high F-values is that more than 80 pct. of the observations have at least one week out of two with zero absence resulting in a perfect correlation between average absence and the absolute value of the first difference for these observations. The model does not take the censoring of absence into account.

Therefore, we also perform the first stage estimation on the sample conditional on some absence. In this regression, we include students with absences in the first and second weeks of school. The results are presented in table 9 and show similar results, with all first-stage estimates significant and relatively high F-values. As before, we know that the high F-values, especially for the instrument using the absolute value of the first difference, are partly driven by many observations, with absence in either week one or week two being close to zero. Further, although the conditioning on positive values of absence may eliminate some of the bias from using a censored variable, it introduces a negative correlation between $\widetilde{a}_{it}$ and η_i in equation 2.4 such that the two are no longer independent and the instrument is therefore not valid. Even though this is the case, we proceed with a description of the results and afterwards briefly discuss a potential avenue of future research to the setup of the panel data instrument.

Table 8: First-stage estimates for the different specifications of the panel data instrument, percent weekly hours absence for the first two weeks used as outcome

	Pct. hours absent (1)	Pct. hours absent (2)	Pct. hours absent (3)	Pct. hours absent (4)
Panel A: All				
Diff. pct. absence	0.232***		0.041***	
	(0.015)		(0.010)	
Diff. pct. absence sq.		0.766***	0.731***	
		(0.028)	(0.028)	
Abs. diff. pct. absence				0.577***
				(0.011)
Observations	5782	5782	5782	5782
F	255.63	740.30	374.71	2836.75
Panel B: With GF1				
Diff. pct. absence	0.201***		0.042**	
	(0.033)		(0.019)	
Diff. pct. absence sq.		0.755***	0.723***	
		(0.041)	(0.040)	
Abs. diff. pct. absence				0.564***
				(0.017)
Observations	1833	1833	1833	1833
F	37.44	337.72	175.97	1089.87
Panel C: Without GF1				
Diff. pct. absence	0.242***		0.041***	
	(0.016)		(0.012)	
Diff. pct. absence sq.		0.764***	0.728***	
		(0.034)	(0.034)	
Abs. diff. pct. absence				0.580***
				(0.013)
Observations	3949	3949	3949	3949
F	227.40	514.78	265.52	2076.57

Notes: The table shows the first-stage estimates for the different specifications of the panel instrument on our endogenous measure of interest, pct. hours absent during the first two weeks. All models contain additional controls (age, immigration status, log distance, and male), year and month fixed effects, and school and education fixed effects. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no proir GF1. F-tests for joint significance are reported for each panel. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 9: First-stage estimates for the different specifications of the panel data instrument, percent weekly hours absence for the first two weeks used as outcome. Sample condtioned on two weeks positive absence

	Pct. hours absent (1)	Pct. hours absent (2)	Pct. hours absent (3)	Pct. hours absent (4)
Panel A: All				
Diff. pct. absence	0.147***		0.0486***	
	(0.0257)		(0.0173)	
Diff. pct. absence sq.		0.613***	0.579***	
		(0.0379)	(0.0367)	
Abs. diff. pct. absence				0.475***
				(0.0240)
Observations	1021	1021	1021	1021
F	32.86	261.4	130.4	391.1
Panel B: GF1				
Diff. pct. absence	0.210***		0.0643**	
	(0.0495)		(0.0316)	
Diff. pct. absence sq.		0.758***	0.697***	
		(0.0791)	(0.0797)	
Abs. diff. pct. absence				0.537***
				(0.0438)
Observations	341	341	341	341
F	18.03	91.76	47.79	150.6
Panel C: without GF1				
Diff. pct. absence	0.117***		0.0354*	
	(0.0295)		(0.0209)	
Diff. pct. absence sq.		0.560***	0.538***	
		(0.0440)	(0.0427)	
Abs. diff. pct. absence				0.444***
				(0.0308)
Observations	680	680	680	680
F	15.62	162.0	81.19	207.9

Notes: The table shows the first-stage estimates for the different specifications of the panel instrument on our endogenous measure of interest, pct. hours absent during the first two weeks, conditional on some absence in both weeks. All models contain additional controls (age, immigration status, log distance, and male), year and month fixed effects, and school and education fixed effects. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no proir GF1. F-tests for joint significance are reported for each panel. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

2.6.3.2 Results: Panel instrument

As mentioned, to support our main results based on the weather measurement instruments, we present the second stage of the IV-estimation, using the panel data instrument for the first two weeks of absence, unconditionally and conditionally on positive absence during both of the two weeks. The results come with the caveat from section 2.6.3.1 that the high F-values are to a large extent driven by a high number of observations with zero absence in either of the the two weeks. Table 10 contains the second stage estimates of the effect of being absent during the first two weeks on the probability of completion based on the four functional forms of $\Delta a_{i2} = \tilde{a}_{i2} - \tilde{a}_{i1}$ from equation 2.4. All columns have completion within seven months as the outcome variable and contain our endogenous variable of interest, absence, measured as percent hours of absence during the first two weeks. We include all controls and month, year, education group, and school fixed effects in all estimations. The bottom panel indicates, for each column, which specification of the first difference in percent hours absent during the first two weeks has been used as the instrument in the associated first-stage regression presented in section 2.6.3.1. We further report first-stage F test values as a measure of the strength of our instruments and p-values for the Anderson-Rubin Wald test, which is robust to weak instruments (Anderson & Rubin, 1949).

In panel A of table 10, which reports results for the full sample, we find that an increase of a 10 percentage point in hours of absence results in a decrease of $7-10.07$ percentage points in the probability of completing, and this roughly corresponds to being absent for one additional day during the first two weeks of school, decreases the probability of completing with $7-10.07$ percentage points. These second-stage estimates are close to the OLS estimate of -1.03, including all control variables and fixed effects; this is most likely caused by the high mechanical correlation between average absence and the first differences in absence when absence in one week is zero. Moving on, we see that panel B, conditional on students with a prior spell in the first basic course, and panel C, conditional on students with no prior spell in the first basic course, show similar results. For panel B, we find that an additional day of absence reduces the probability of completing with $5-10$ percentage points, and for panel C, we find that an additional day of absence reduces the probability of completing with $7.6-10$ percentage points. Again all estimates in panels B and C are highly significant, with high F values and low Anderson-Rubin p-values.

We also perform the second-stage regressions for the absence during the first two weeks, conditional on some absence in both weeks. We present the results in Table 11. Unfortunately, the large drop in the number of observations results in a large loss of power, such that none of the estimated coefficients are significant, and their magnitude is much smaller.

Table 10: Effect of absence during the first two weeks on probability of graduation: 2SLS with the panel instrument as IV for absence

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)
Panel C: All				
Pct. hours absent	-1.068***	-0.701***	-0.720***	-0.861***
	(0.172)	(0.071)	(0.073)	(0.059)
Observations	5782	5782	5782	5782
First-stage F	255.63	740.30	374.71	2836.75
AR p-value	0.000	0.000	0.000	0.000
Panel B: With GF1				
Pct. hours absent	-1.022***	-0.495***	-0.521***	-0.746***
	(0.311)	(0.161)	(0.162)	(0.132)
Observations	1833	1833	1833	1833
First-stage F	37.44	337.72	175.97	1089.87
AR p-value	0.007	0.008	0.005	0.000
Panel A: Without GF1				
Pct. hours absent	-1.035***	-0.757***	-0.772***	-0.889***
	(0.186)	(0.088)	(0.090)	(0.080)
Observations	3949	3949	3949	3949
First-stage F	227.40	514.78	265.52	2076.57
AR p-value	0.000	0.000	0.000	0.000
$\Delta absence$	Yes	No	Yes	No
$(\Delta absence)^2$	No	Yes	Yes	No
$\|\Delta absence\|$	No	No	No	Yes

Notes: The dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are the first difference in weekly absence, the first difference in weekly absence squared, and the absolute difference in weekly absence. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Effect of absence during the first two weeks on probability of graduation. Sample condtioned on two weeks positive absence: 2SLS with the panel instrument as IV for absence

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)
Panel A: All				
Pct. hours absent	-0.400	-0.174	-0.192	-0.310*
	(0.370)	(0.174)	(0.172)	(0.160)
Observations	1021	1021	1021	1021
First-stage F	32.86	261.4	130.4	391.1
AR p-value	0.298	0.337	0.475	0.0631
Panel B: GF1				
Pct. hours absent	-0.711	-0.262	-0.308	-0.409
	(0.565)	(0.325)	(0.321)	(0.310)
Observations	341	341	341	341
First-stage F	18.03	91.76	47.79	150.6
AR p-value	0.237	0.466	0.489	0.226
Panel C: Without GF1				
Pct. hours absent	-0.274	-0.177	-0.182	-0.321*
	(0.530)	(0.210)	(0.209)	(0.193)
Observations	680	680	680	680
First-stage F	15.62	162.0	81.19	207.9
AR p-value	0.627	0.427	0.704	0.118
$\Delta absence$	Yes	No	Yes	No
$(\Delta absence)^2$	No	Yes	Yes	No
$\lvert\Delta absence\rvert$	No	No	No	Yes

The dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are the first difference in weekly absence, the first difference in weekly absence squared, and the absolute difference in weekly absence. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

2.7 Discussion

As mentioned above, the panel data instrument suffers from being censored in week 1 and week 2 and therefore gives improbable high correlation between the average absence during week 1 and 2 and the first difference in absence from the two weeks. We try to abate this issue by only analyzing first difference in absence for students with positive absence in both weeks however, this instead introduces a negative correlation between $\widetilde{a}_{it}$ and η_i, such that the two are no longer independent and the instrument is therefore not valid. In future work, we plan to include absence during the last year of compulsory school (grade 9). This way, we can create an instrument that is the average absence during the first two weeks subtracted by the average absence during the 9th grade. We would need to condition on positive absence during the first two weeks of vocational school, but would most likely not need to do any conditioning on 9th grade absence, if all students have at least one hour of absence during the year. This way we would avoid the censoring and break the negative correlation between between $\widetilde{a}_{it}$ and η_i.

Another planned extension is to wait for Statistics Denmark to release new population data on vocational school absence. We believe this data will be released some time in 2023. With population data, instead of data from 8 schools, we will have more variation in starting dates and more geographical variation that can hopefully strengthen our weather instrument and give more robust second stage estimate from this instruments.

2.8 Conclusion

To conclude, we find that absence during the first weeks of the second basic course can account for many of the students who drop out. Our preferred estimate entails that one day of additional absence for students in the second basic course with no prior spell in the first basic course are 19.5 percentage points less likely to graduate the second basic course within seven months. We find no significant results for the full sample or the sample of students with a prior spell in the first basic course. Our estimated coefficient is larger in magnitude than the OLS estimate.

We find that weather is stronger correlated with the groups of students who we expect to be more on the margin between attending school or not when the weather is bad, namely the students who do not live with their parents and student with no previous labor market attachment or school enrollment. This indicates that these are the group of students driving our first stage results and the identification of our results.

We check the robustness of our results with a second approach to identify the causal effect of absence on the probability of graduating, namely a new panel instrument. We find estimates which are close to the OLS estimates and lower than the results using weather measurements as instruments. These results are driven by the fact that our panel data instruments are generated from weekly observations of absence, which, to a high degree, are bottom censored. In future research, we aim to explore other suitable absence measures for students who do not suffer from as severe bottom censoring.

References

Anderson, T. W., & Rubin, H. (1949). Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *Annals of Mathematical Statistics*, *20*, 46–63.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*, 29–51.

Aucejo, E. M., & Romano, T. F. (2016). Assessing the effect of school days and absences on test score performance. *Economics of Education Review*, *55*, 70–87. https://doi.org/10.1016/j.econedurev.2016.08.007

Auffhammer, M., Hsiang, S. M., Schlenker, W., & Sobel, A. (2013). Using Weather Data and Climate Model Output in Economic Analyses of Climate Change. *NBER Working Paper Series*, *No. 19087*.

Bertrand, M., Mogstad, M., & Mountjoy, J. (2021). Improving educational pathways to social mobility: Evidence from norway's reform 94. *Journal of Labor Economics*, *39*(4), 965–1010.

Bingley, P., Heinesen, E., Krassel, K. F., & Kristensen, N. (2018). The Timing of Instruction Time: Accumulated Hours, Timing and Pupil Achievement. *IZA Discussion Papers*, *No. 11807*.

Carlsson, M., Dahl, G. B., Öckert, B., & Rooth, D.-O. (2015). The Effect of Schooling on Cognitive Skills. *The Review of Economics and Statistics*, *97*(3), 533–547. https://doi.org/10.1162/REST

Cattan, S., Kamhöfer, D. A., Karlsson, M., & Nilsson, T. (2023). The Long-term Effects of Student Absence : Evidence from Sweden. *The Economic Journal*, *133*, 888–903.

Confederation of Danish Employers. (2023). Frafald på erhvervsuddannelserne sker oftest på grundforløbet [Accessed: 2023-04-18].

Craig, A. C., & Martin, D. C. (2019). *Discipline Reform, School Culture, and Student Achievement Ashley* [Working Paper].

Currie, J., Hanushek, E. A., Kahn, E. M., Neidell, M., & Rivkin, S. G. (2009). Does Pollution Increase School Absences? *The Review of Economics and Statistics*, *91*(4), 682–694.

Dell, M., Jones, B. F., & Olken, B. A. (2014). What do we learn from the weather? The new climate-economy literature. *Journal of Economic Literature*, *52*(3), 740–798. https://doi.org/10.1257/jel.52.3.740

Dominguez. Patricio & Ruffini, K. (2018). Long-term gains from longer school days. *IDB Working Paper Series*, *No. 1120*. http://irle.berkeley.edu/working-papers

Eichhorst, W., Rodríguez-Planas, N., Schmidl, R., & Zimermann, K. F. (2015). A road map to vocational education and training in industrialized countries. *Industrial and Labor Relations Review*, *68*(2), 314–337. https://doi.org/10.1177/0019793914564963

Fitzpatrick, M. D., Grissmer, D., & Hastedt, S. (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review*, *30*(2), 269–279. https://doi.org/10.1016/j.econedurev.2010.09.004

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, *11*(2), 125–149. https://doi.org/10.1162/EDFP_a_00180

Goodman, J. S. (2014). Flaking out: Student absences and snow days as disruptions of instruction time. *NBER Working Paper Series*, *No. 20221*. https://doi.org/10.3386/w20221

Gottfried, M. A. (2009). Excused versus unexcused: How student absences in elementary school affect academic achievement. *Educational Evaluation and Policy Analysis*, *31*(4), 392–415. https://doi.org/10.3102/0162373709342467

Gottfried, M. A. (2010). Evaluating the relationship between student attendance and achievement in urban elementary and middle schools: An instrumental variables approach. *American Educational Research Journal*, *47*(2), 434–465. https://doi.org/10.3102/0002831209350494

Groes, F. N., Madsen, E., & Sandoy, T. M. (2021). *Completion from vocational educations: A register based analysis* (tech. rep.). Copenhagen Business School.

Groppo, V., & Kraehnert, K. (2017). The impact of extreme weather events on education. *Journal of Population Economics*, *30*(2), 433–472. https://doi.org/10.1007/s00148-016-0628-6

Hampf, F., & Woessmann, L. (2017). Vocational vs. general education and employment over the life cycle: New evidence from PIAAC. *CESifo Economic Studies*, *63*(3), 255–269. https://doi.org/10.1093/cesifo/ifx012

Hansen, B. (2011). School Year Length and Student Performance: Quasi-Experimental Evidence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2269846

Hanushek, E. A. (2012). Dual Education: Europe's Secret Recipe? *CESifo Forum*, *13*(3), 29–34.

Hanushek, E. A., Schwerdt, G., Woessmann, L., & Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the lifecycle. *Journal of Human Resources*, *52*(1), 48–87. https://doi.org/10.3368/jhr.52.1.0415-7074R

Heissel, J. A., & Norris, S. (2018). Rise and shine: The effect of school start times on academic performance from childhood through puberty. *Journal of Human Resources*, *53*(4), 957–992. https://doi.org/10.3368/jhr.53.4.0815-7346R1

Kristensen, N., Jensen, V. M., & Krassel, K. F. (2020). *Panelanalyse af bekymrende skolefravær* (tech. rep.). VIVE.

Lavy, V. (2015). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. *Economic Journal*, *125*(588), F397–F424. https://doi.org/10.1111/ecoj.12233

Lavy, V. (2020). Expanding School Resources and Increasing Time on Task: Effects on Students' Academic and Noncognitive Outcomes. *Journal of the European Economic Association*, *18*(1), 232–265. https://doi.org/10.1093/jeea/jvy054

Liu, J., Lee, M., & Gershenson, S. (2019). The Short-and Long-Run Impacts of Secondary School Absences. *IZA Discussion Papers*, *No. 12613*.

Liu, J., & Loeb, S. (2019). Engaging Teachers: Measuring the Impact of Teachers on Student Attendance in Secondary School. *Journal of Human Resources*, *56*(2), 343–379. https://doi.org/10.3368/jhr.56.2.1216-8430r3

Marcotte, D. E. (2007). Schooling and test scores: A mother-natural experiment. *Economics of Education Review*, *26*(5), 629–640. https://doi.org/10.1016/j.econedurev.2006.08.001

Marcotte, D. E., & Hemelt, S. W. (2008). Unscheduled School Closings and Student Performance. *Education Finance and Policy*, *3*(3), 316–338. https://doi.org/10.1162/edfp.2008.3.3.316

Mellon, J. (2020). Rain, Rain, Go Away: 137 Potential Exclusion-Restriction Violations for Studies Using Weather as an Instrumental Variable. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3715610

Park, R. J., Goodman, J., Hurwitz, M., & Smith, J. (2020). Heat and learning. *American Economic Journal: Economic Policy*, *12*(2), 306–339. https://doi.org/10.1257/POL.20180612

Rivkin, S. G., & Schiman, J. C. (2015). Instruction Time, Classroom Quality, and Academic Achievement. *Economic Journal*, *125*(588), F425–F448. https://doi.org/10.1111/ecoj.12315

Robinson, C. D., Lee, M. G., Dearing, E., & Rogers, T. (2018). Reducing Student Absenteeism in the Early Grades by Targeting Parental Beliefs. *American Educational Research Journal*, *55*(6), 1163–1192. https://doi.org/10.3102/0002831218772274

Rogers, T., & Feller, A. (2016). Discouraged by Peer Excellence: Exposure to Exemplary Peer Performance Causes Quitting. *Psychological Science*, *27*(3), 365–374. https://doi.org/10.1177/0956797615623770

Sarsons, H. (2015). Rainfall and conflict: A cautionary tale. *Journal of Development Economics*, *115*, 62–72. https://doi.org/10.1016/j.jdeveco.2014.12.007

Silliman, M., & Virtanen, H. (2022). Labor market returns to vocational secondary education. *American Economic Journal: Applied Economics*, *14*(1), 197–224.

Stratton, L. S., Gupta, N. D., Reimer, D., & Holm, A. (2017). *Modeling Enrollment in and Completion of Vocational Education: The role of cognitive and non-cognitive skills by program type* (No. 2017-2) [Working Paper], Aarhus BSS.

The Danish Ministry of Education. (2018). Karakterer fra folkeskolens afgangseksamen [Accessed: 2023-04-17].

The Danish Ministry of Education. (2019). Vocational education and training in Denmark [Accessed: 2023-04-18].

Thejill, P., Boberg, F., Schmith, T., Christiansen, B., Christensen, O. B., Madsen, M. S., Su, J., Andree, E., Olsen, S., Langen, P. L., Madsen, K. S., & Pedersen, R. A. (2020). *Methods used in the Danish Climate Atlas* (tech. rep.). The Danish Meteorological Institute. Copenhagen.

Tran, L., & Gershenson, S. (2018). Experimental Estimates of the Student Attendance Production Function. *IZA Discussion Paper*, *11911*.

Zimmer, D. M. (2019). *Missing School to Visit the Doctor? Analysis Using a Copula-Based Endogenous Switching Regressions Model* [Working Paper].

Appendices

2.A Additional tables and figures

Table 2.A.1: The Effect of pct. days with some absence during the first two weeks on graduation: OLS results

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)	Graduated (5)	Graduated (6)	Graduated (7)
Panel A: All							
Pct. days absence	-0.997***	-0.885***	-0.875***	-0.996***	-0.981***	-0.979***	-0.867***
	(0.048)	(0.050)	(0.050)	(0.049)	(0.053)	(0.053)	(0.050)
Observations	5782	5782	5782	5782	5782	5782	5782
R-squared	0.059	0.129	0.139	0.060	0.071	0.072	0.157
Panel B: With GF1							
Pct. days absence	-1.004***	-0.882***	-0.881***	-1.011***	-1.010***	-1.020***	-0.914***
	(0.137)	(0.110)	(0.112)	(0.142)	(0.142)	(0.149)	(0.114)
Observations	1833	1833	1833	1833	1833	1833	1833
R-squared	0.042	0.119	0.131	0.044	0.064	0.065	0.167
Panel C: Without GF1							
Pct. days absence	-0.992***	-0.869***	-0.858***	-0.985***	-0.949***	-0.942***	-0.834***
	(0.058)	(0.066)	(0.066)	(0.059)	(0.061)	(0.061)	(0.067)
Observations	3949	3949	3949	3949	3949	3949	3949
R-squared	0.065	0.143	0.154	0.068	0.081	0.084	0.172
Student controls	No	Yes	Yes	No	No	No	Yes
Parent controls	No	No	Yes	No	No	No	Yes
Year/Month FE	No	No	No	Yes	No	Yes	Yes
School/Education FE	No	No	No	No	Yes	Yes	Yes

Notes: The table shows the raw OLS estimates for pct. days with some absence during the first two weeks on graduation (0/1). The bottom panel indicates if student controls (age, immigration status, log distance, male, 9th grade danish score, 9th grade math score, and labour market attachment prior to enrolment), parent controls (highest completed parental education, mothers and fathers labour market attachment), year and month fixed effects, and school and education fixed effects are included in the models in the respective column. Panel A uses the full sample, Panel B only uses students with a prior GF1, and Panel C only uses students with no prior GF1. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.2: Effect of absence on probability of graduation: 2SLS with weather as IV for absence (2 weeks)

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)
Panel A: All				
Pct. hours absent	-3.644	-1.068	-1.898*	-1.841**
	(3.085)	(1.438)	(1.104)	(0.911)
Observations	5782	5782	5782	5782
First-stage F	2.42	6.51	3.82	3.06
AR p-value	0.050	0.493	0.109	0.159
Panel B: With GF1				
Pct. hours absent	-1.952**	-1.059	-1.404**	-1.357**
	(0.928)	(0.817)	(0.636)	(0.611)
Observations	3949	3949	3949	3949
First-stage F	9.76	19.51	12.59	11.27
AR p-value	0.004	0.205	0.018	0.045
Precipitation $> 3mm$	Yes	No	Yes	Yes
Wind $> 11ms$	No	Yes	Yes	Yes
Precipitation×Wind	No	No	No	Yes

The dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 2.A.3: Precipitation specification table, first-stage (2 weeks), without GF1

	1 mm	2 mm	3 mm	4 mm	5 mm	6 mm	7 mm	8 mm	9 mm
Precipitation	0.065***	0.075***	0.082***	0.075**	0.087**	0.047	0.108***	0.129***	0.101***
	(0.013)	(0.022)	(0.026)	(0.032)	(0.037)	(0.029)	(0.040)	(0.041)	(0.038)
F	25.00	11.73	9.76	5.44	5.53	2.54	7.41	9.89	7.18

Notes: The reported estimates are the first-stage estimates from a regression with pct. hours of absence during the first two weeks with precipitation as instrument, using from 1 to 9 mm of precipitation as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.4: Precipitation specification table, 2SLS (2 weeks), without GF1

	1 mm	2 mm	3 mm	4 mm	5 mm	6 mm	7 mm	8 mm	9 mm
Pct. hours absent	-1.687*	-0.942	-1.952**	-1.705*	-2.193*	-5.900	-1.615	-1.652	-1.718
	(0.906)	(0.809)	(0.928)	(1.009)	(1.243)	(4.443)	(1.228)	(1.106)	(1.449)
Observations	3949	3949	3949	3949	3949	3949	3949	3949	3949
F	25.00	11.73	9.76	5.44	5.53	2.54	7.41	9.89	7.18
AR p-value	0.039	0.226	0.004	0.026	0.048	0.019	0.138	0.139	0.225

Notes: The reported estimates are the second-stage estimates from a regression with pct. hours of absence as endogenous variable while the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The instrument used is precipitation, using from 1 to 9 mm of precipitation as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance from the first-stage are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.5: Wind specification table, first-stage (2 weeks), without GF1

	10 ms	11 ms	12 ms	13 ms	14 ms	15 ms	16 ms	17 ms	18 ms
Wind	0.071***	0.090***	0.082***	0.082***	0.084**	0.108***	0.119**	0.142	0.166*
	(0.022)	(0.020)	(0.021)	(0.019)	(0.034)	(0.028)	(0.052)	(0.090)	(0.096)
F	10.11	19.51	15.06	19.33	6.11	14.74	5.33	2.47	2.99

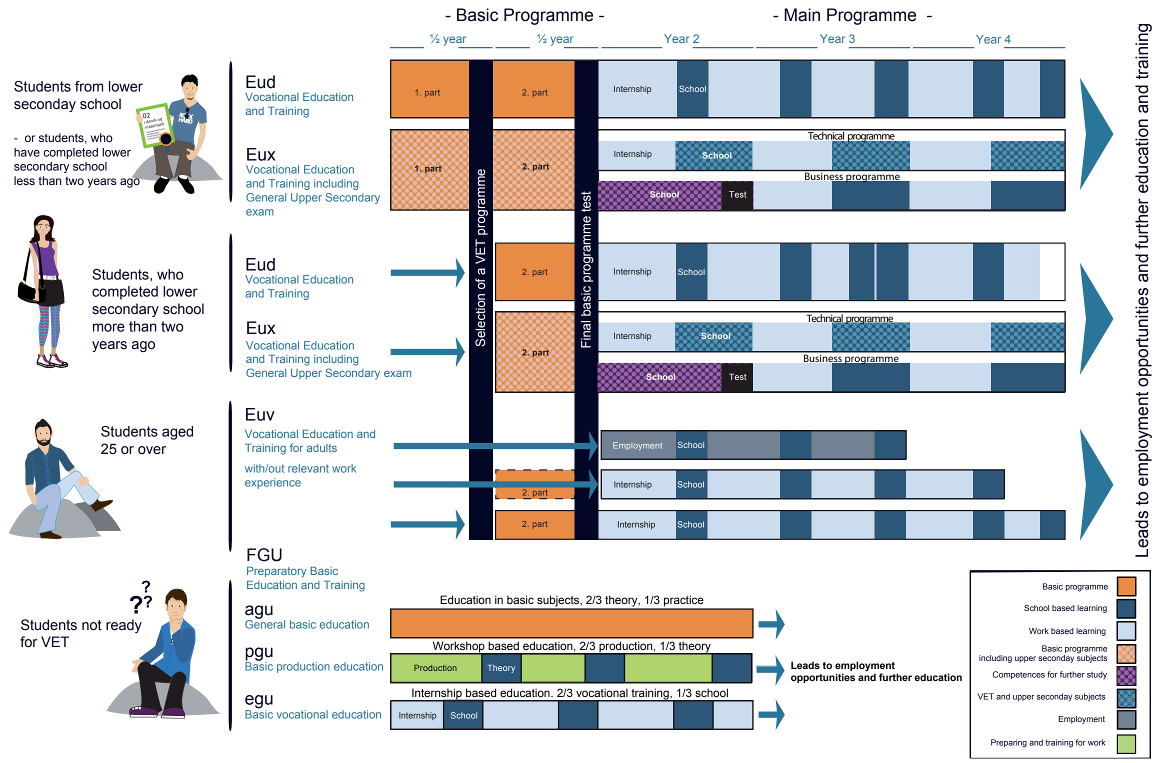
Notes: The reported estimates are the first-stage estimates from a regression with pct. hours of absence during the first two weeks with high wind as instrument, using from 10 to 18 ms of wind speed as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.6: Wind specification table, 2SLS (2 weeks), without GF1

	10 ms	11 ms	12 ms	13 ms	14 ms	15 ms	16 ms	17 ms	18 ms
Pct. hours absent	-1.828*	-1.059	-1.614*	-1.346*	0.316	-0.534	-1.612	0.960	1.755
	(0.945)	(0.817)	(0.850)	(0.749)	(1.039)	(0.862)	(1.269)	(1.747)	(1.404)
Observations	3949	3949	3949	3949	3949	3949	3949	3949	3949
F	10.11	19.51	15.06	19.33	6.11	14.74	5.33	2.47	2.99
AR p-value	0.025	0.205	0.048	0.075	0.758	0.549	0.221	0.601	0.193

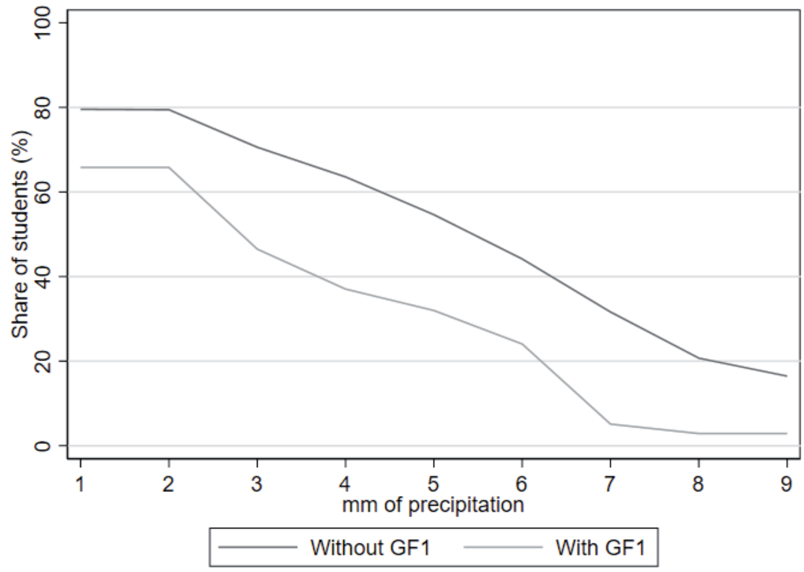
Notes: The reported estimates are the second-stage estimates from a regression with pct. hours of absence as endogenous variable while the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1). The instrument used is high wind speed, using from 10 to 18 ms of wind speed as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance from the first-stage are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Figure 2.A.1: Pathways through the Danish VET system



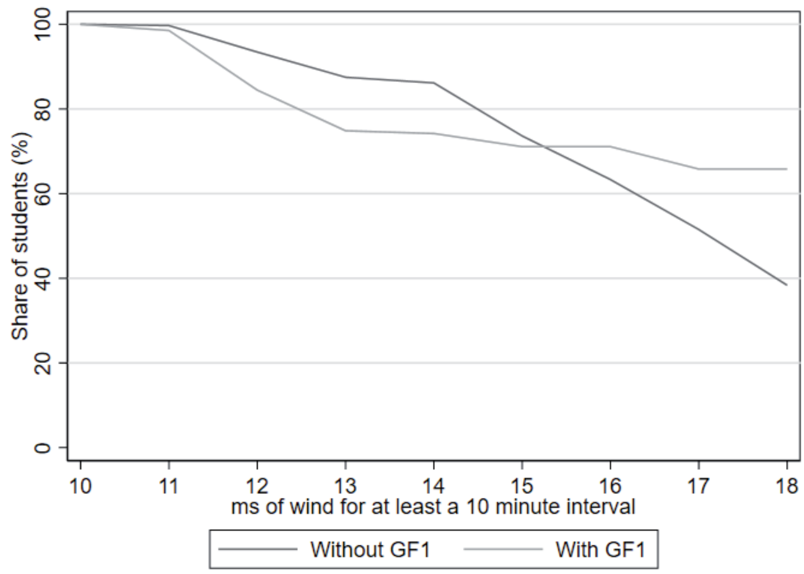
Note: The figure is taken from the web-page of the Danish Ministry of Children and Education, where a description of the Danish VET system can also be found (The Danish Ministry of Education, 2019).

Figure 2.A.2: Share of students who have experienced at least one day with precipitation above a given number of mm



Note: The figure has the share of students who have experienced at least one day with a given number of mm or precipitation or more on the second axis and mm of precipitation on the first axis.

Figure 2.A.3: Share of students who have experienced at least one day with wind above a given speed of ms



Note: The figure has the share of students who have experienced at least one day with wind speed of a given ms or more on the second axis and ms of wind on the first axis.

Table 2.A.7: Interaction of precipitation and wind specification matrix table, 2SLS (2 weeks), without GF1

	1 mm	2 mm	3 mm	4 mm	5 mm	6 mm	7 mm	8 mm	9 mm
10 ms	-1.741**	-1.390*	-1.888**	-1.792**	-1.940**	-1.920**	-1.770*	-1.778**	-1.815*
	(0.826)	(0.760)	(0.806)	(0.853)	(0.879)	(0.958)	(0.922)	(0.885)	(0.942)
F	11.54	7.46	7.00	5.63	5.85	5.33	6.50	7.25	5.44
AR p-value	0.042	0.072	0.010	0.031	0.052	0.021	0.069	0.062	0.075
11 ms	-1.352*	-1.013	-1.404**	-1.210*	-1.340*	-1.104	-1.130	-1.138	-1.068
	(0.725)	(0.664)	(0.636)	(0.708)	(0.736)	(0.817)	(0.825)	(0.819)	(0.821)
F	17.20	12.54	12.59	10.00	10.51	9.95	10.97	11.12	9.90
AR p-value	0.113	0.303	0.018	0.085	0.110	0.062	0.309	0.291	0.391
12 ms	-1.655**	-1.309*	-1.767**	-1.641**	-1.794**	-1.799**	-1.614*	-1.623*	-1.626*
	(0.795)	(0.719)	(0.694)	(0.736)	(0.750)	(0.844)	(0.858)	(0.830)	(0.865)
F	15.58	11.03	11.53	8.27	8.90	7.93	9.16	9.35	7.89
AR p-value	0.083	0.129	0.008	0.043	0.044	0.024	0.134	0.120	0.139
13 ms	-1.605**	-1.101	-1.691***	-1.489**	-1.700**	-1.564**	-1.446*	-1.465**	-1.439*
	(0.786)	(0.671)	(0.654)	(0.676)	(0.731)	(0.777)	(0.788)	(0.712)	(0.746)
F	17.36	12.14	13.26	10.15	10.10	10.58	10.68	14.75	11.97
AR p-value	0.093	0.177	0.011	0.074	0.079	0.054	0.160	0.137	0.179
14 ms	-1.307*	-0.539	-1.113*	-0.625	-0.903	-0.375	-0.544	-0.573	-0.325
	(0.771)	(0.649)	(0.607)	(0.684)	(0.712)	(0.908)	(0.795)	(0.792)	(0.887)
F	13.10	7.26	7.39	4.18	5.13	5.20	6.89	9.73	6.22
AR p-value	0.046	0.411	0.014	0.063	0.116	0.050	0.267	0.267	0.384
15 ms	-1.397*	-0.778	-1.333**	-1.012	-1.235	-0.934	-0.962	-0.989	-0.878
	(0.780)	(0.665)	(0.665)	(0.694)	(0.782)	(0.899)	(0.768)	(0.743)	(0.800)
F	14.82	9.90	11.44	8.26	8.91	9.38	11.46	15.81	11.75
AR p-value	0.097	0.463	0.015	0.081	0.140	0.064	0.316	0.301	0.430
16 ms	-1.684*	-1.049	-1.867**	-1.670*	-1.991**	-2.436*	-1.614*	-1.636*	-1.662*
	(0.897)	(0.777)	(0.835)	(0.897)	(0.995)	(1.346)	(0.962)	(0.893)	(1.007)
F	12.97	7.15	7.52	4.52	5.46	5.51	8.06	11.50	7.72
AR p-value	0.115	0.329	0.013	0.073	0.125	0.066	0.231	0.174	0.242
17 ms	-1.739*	-0.897	-1.564**	-1.180	-1.647*	-2.007	-1.115	-1.032	-0.858
	(0.913)	(0.801)	(0.763)	(0.768)	(0.985)	(1.736)	(1.116)	(1.035)	(1.238)
F	13.08	6.84	7.01	5.39	5.56	4.17	5.42	11.63	7.50
AR p-value	0.027	0.301	0.017	0.079	0.114	0.052	0.235	0.270	0.385
18 ms	-1.786*	-0.937	-1.614**	-1.208	-1.702*	-1.604	-0.996	-1.049	-0.769
	(0.927)	(0.807)	(0.779)	(0.787)	(1.023)	(1.674)	(1.021)	(0.974)	(1.152)
F	13.82	7.46	7.41	5.71	5.59	3.59	5.35	8.01	6.29
AR p-value	0.017	0.135	0.014	0.056	0.063	0.034	0.137	0.125	0.180

Notes: The reported estimates are the first-stage estimates from a regression with pct. hours of absence during the first two weeks with the precipitation instrument interacted with the high wind instrument, using from 1 to 9 mm of precipitation (columns) and 10 to 18 ms of wind speed (rows) as cutoffs for the instrument. Standard errors clustered on the school by start date level are reported in parentheses and F-tests for joint significance are reported as well. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.8: First-stage estimates for the different specifications of weather instruments, percent days absent used as outcome (2 weeks), without GF1

	Pct. days absent (1)	Pct. days absent (2)	Pct. days absent (3)	Pct. days absent (4)
Pct. days with precipitation>3mm	0.073**		0.052	0.014
	(0.036)		(0.038)	(0.036)
Pct. days with wind>11ms		0.066***	0.046**	0.025
		(0.023)	(0.022)	(0.024)
Precipitation X wind				0.179
				(0.111)
Observations	3949	3949	3949	3949
F	4.05	8.29	5.60	4.51

Notes: The table shows the first-stage estimates for the different specifications of the weather instruments on our endogenous measure of interest, pct. days absent during the first three weeks. All models contain additional controls (age, immigration status, log distance, and male), year and month fixed effects, and school and education fixed effects. F-tests for joint significance are reported for each panel. Standard errors clustered on the school by start date level are reported in parentheses. Significance levels are reported as *** p<0.01, ** $p < 0.05$, * $p < 0.1$.

Table 2.A.9: Effect of absence on probability of graduation: 2SLS with weather as IV for absence (2 weeks)

	Graduated (1)	Graduated (2)	Graduated (3)	Graduated (4)
Pct. days absent	-2.195*	-1.481*	-1.848**	-1.809**
	(1.120)	(0.871)	(0.828)	(0.793)
Observations	3949	3949	3949	3949
F	4.05	8.29	5.60	4.51
AR p-value	0.001	0.079	0.003	0.004
Precipitation > $3mm$	Yes	No	Yes	Yes
Wind > $11ms$	No	Yes	Yes	Yes
Precipitation×Wind	No	No	No	Yes

The dependent variable is a dummy indicating if the student has graduated within 7 months (0/1) and the endogenous variable is pct. days absent during the first two weeks. The bottom panel contains indicators for which instruments have been used in the first-stage estimation. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 2.A.10: Heterogeneity in terms of living with parents: 2SLS with weather as IV for absence (2 weeks)

	Does not live with parents		Lives with parents	
	(1)	(2)	(3)	(4)
Second stage estimates				
Pct. hours absent	-1.765**	-1.832	-2.386	-0.061
	(0.850)	(1.173)	(3.405)	(1.178)
OLS estimates				
Pct. hours absent	-0.899***		-0.847***	
	(0.077)		(0.089)	
First stage estimates				
Pct. days with precipitation>3mm	0.093***		0.021	
	(0.019)		(0.019)	
Pct. days with wind>11ms		0.094***		0.073***
		(0.031)		(0.027)
Observations	2287	2287	1662	1662
F	24.11	9.03	1.16	7.25
AR p-value	0.023	0.110	0.437	0.960
Precipitation $> 3mm$	Yes	No	Yes	No
Wind $> 11ms$	No	Yes	No	Yes

All columns only include students without the first basic course. The first two columns only include students who do not live with their parents, while the third and fourth columns include students who live with their parents. The top panel is the second-stage estimates from a regression where the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1) and the endogenous variable is pct. hours absent during the first two weeks using the instruments marked in the bottom panel. The second panel shows the OLS estimates from running the same regression without instruments. The third panel reports the first-stage estimates using pct. hours absent as dependent variable and the instruments marked in the bottom panel as explanatory varibles. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 2.A.11: Heterogeneity in terms of a prior stable attachment: 2SLS with weather as IV for absence (2 weeks)

	Work or in school		No stable attachment	
	(1)	(2)	(3)	(4)
Second stage estimates				
Pct. hours absent	-2.253**	-1.351	-1.334	-0.758
	(0.915)	(0.830)	(1.112)	(1.282)
OLS estimates				
Pct. hours absent	-0.760***	-0.760***	-1.118***	-1.118***
	(0.075)	(0.075)	(0.077)	(0.077)
First stage estimates				
Pct. days with precipitation>3mm	0.057***		0.096***	
	(0.011)		(0.024)	
Pct. days with wind>11ms		0.081***		0.115***
		(0.019)		(0.044)
Observations	2912	2912	1037	1037
F	27.10	18.89	15.56	6.71
AR p-value	0.005	0.117	0.245	0.584
Precipitation $> 3mm$	Yes	No	Yes	No
Wind $> 11ms$	No	Yes	No	Yes

All columns only include students without the first basic course. The first two columns only include students who where working or in school before their enrolment, while the third and fourth columns include students without a stable attachment before their enrolment. The top panel is the second-stage estimates from a regression where the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1) and the endogenous variable is pct. hours absent during the first two weeks using the instruments marked in the bottom panel. The second panel shows the OLS estimates from running the same regression without instruments. The third panel reports the first-stage estimates using pct. hours absent as dependent variable and the instruments marked in the bottom panel as explanatory varibles. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

Table 2.A.12: Heterogeneity in terms of grades: 2SLS with weather as IV for absence (2 weeks)

	High grade		Low grade	
	(1)	(2)	(3)	(4)
Second stage estimates				
Pct. hours absent	-3.163**	-4.241**	-0.918	1.715
	(1.504)	(2.091)	(1.218)	(1.436)
OLS estimates				
Pct. hours absent	-0.855***		-0.838***	
	(0.073)		(0.089)	
First stage estimates				
Pct. days with precipitation>3mm	0.049**		0.069***	
	(0.019)		(0.027)	
Pct. days with wind>11ms		0.055**		0.117***
		(0.025)		(0.035)
Observations	1942	1942	1297	1297
F	6.33	4.87	6.66	11.09
AR p-value	0.007	0.007	0.423	0.164
Precipitation $> 3mm$	Yes	No	Yes	No
Wind $> 11ms$	No	Yes	No	Yes

All columns only include students without the first basic course. The first two columns only include students who had a 9th grade Math grade of 4 or below, while the third and fourth columns include students who had a higher 9th grade Math grade. The top panel is the second-stage estimates from a regression where the dependent variable is a dummy indicating if the student has graduated within 7 months (0/1) and the endogenous variable is pct. hours absent during the first two weeks using the instruments marked in the bottom panel. The second panel shows the OLS estimates from running the same regression without instruments. The third panel reports the first-stage estimates using pct. hours absent as dependent variable and the instruments marked in the bottom panel as explanatory varibles. The instruments used are pct. of days with over 3 mm of precipitation within the first two weeks, pct. days with average windspeed above 11 ms for a 10 min. interval, and the interaction of these two instruments. All models further include the log distance from students parish to the schools parish, age dummies, a dummy for males, dummies for first and second generation immigrants, and year, month, school, and education fixed effects. Standard errors clustered on the school by start date level are reported in parentheses and the reported F-values are from the first-stage. AR p-values are from the Anderson-Rubin wald test. Significance levels are reported as *** p<0.01, ** p<0.05, * p<0.1.

CHAPTER 3

University Peers and Labour Market Gender Gaps

with Alexander Fischer, Andrei Gorshkov, and Jeanette Walldorf

Abstract

This paper investigates how university peers affect the divergence in career trajectories of male and female students to top-earning jobs in the labor market. We use university records covering 21 cohorts of incoming students randomly assigned to peer groups in a large Business Economics program in Denmark and merge it with high-quality administrative register data on students' careers. This setup allows us to estimate the effect of peer ability on students' labour market outcomes many years post matriculation.

We find that male students assigned to high-achieving peers are largely unaffected. On the other hand, we find that female students assigned to peers of higher ability suffer severe earnings losses, have weaker labour market attachment, are less likely to work in positions with management responsibilities, and are less likely to reach the top of the earnings distribution.

The negative effect on female careers is more pronounced for high-ability female students and driven by exposure to high-achieving male peers. To explore potential reasons for the negative effect, we investigate how peer ability affects educational outcomes and family investment.

3.1 Introduction

Over the last decades, developed countries have made significant progress in improving the labour market positions of women (Goldin, 2014, Olivetti and Petrongolo, 2016, Blau and Kahn, 2017, Petrongolo and Ronchi, 2020). Nevertheless, women remain strongly underrepresented in high-paying managerial positions (Bertrand, 2018). In S&P 500 companies, while representing around 45% of the workforce, females hold only 5.8% of

CEO positions, and only 11% of top earners and 21.2% of board members are female (Catalyst, 2020). In developed economies, typically less than one-third of the members of the highest decile of the income distribution are females, and the number reduces to less than one fifth for the top 1% (Atkinson et al., 2018).

A growing literature on peer effects suggests that social interactions in educational institutions determine educational choices and the academic success of students (Feld and Zölitz, 2017, Brenøe and Zölitz, 2020). Moreover, peers seem to affect male and female students differently (Feld and Zölitz, 2018, Mouganie and Wang, 2020, Fischer, 2017). Given that peers influence educational choices, do they also matter for the career outcomes of potential top earners? Furthermore, given that peers have differential effects on male and female students, do peers in university contribute to the gender difference in top-earning jobs?

Answering such questions poses several challenges to both data and research design. First, data on peer groups and students' careers must be available. Most importantly, the peer group assignment needs to be uncorrelated with unobservable factors that also affect career outcomes.

To address these challenges, we link information on 21 cohorts of Business Economics students at Copenhagen Business School to Danish register data, which provides rich information on the students' careers. Furthermore, we can identify the causal effect of peer ability on male and female careers by utilizing a unique feature of our data: before matriculation, the students in our sample were randomly assigned to peer groups of around 35 peers.

Denmark, among other Scandinavian countries, is often considered a frontrunner in gender equality (number 14 according to the Global Gender Gap Report 2020 country ranking in "overall gender equality"). At the same time, Denmark scores low in the share of women working as lawmakers, senior officials, and managers (as low as 102nd in the same country ranking). Therefore, Denmark offers a compelling case for studying if peers contribute to gender differences at the top of the income distribution.[1]

Our findings suggest that peer ability composition has a lasting effect on labour market outcomes: while we find little evidence that the career success of males is determined by peer ability, female students' careers are severely negatively affected by exposure to peers of higher average ability. Women exposed to high-achieving peers experience a substantial earnings penalty: increasing their peers' average ability by one standard deviation decreases females' earnings by around 9%. Furthermore, exposure to higher-ability peers decreases females' propensities to work in jobs with management responsibilities and to reach the top of the income distribution. The career penalty for female students is driven by exposure to high-achieving male peers and female students with high ability are more adversely affected than female low ability students.

We explore potential reasons that might explain the negative impact of higher-ability peers on female careers.First, we investigate if peers adversely affect female students'

[1]Gallen et al. (2019) investigate the dynamics of the gender gap in Denmark over the last 30 years. For a discussion of gender gaps at the top of the career ladder in Denmark, see Smith et al. (2011), and Datta Gupta et al. (2006). Albrecht et al. (2003) investigate gender gaps in another Scandinavian country, Sweden.

human capital formation. Female students exposed to peers of higher ability graduate with significantly worse GPA's than their graduating cohort peers. Further, they are also less likely to continue their education and graduate with a master degree.

Second, we inspect if changing family formation behavior explains female students' adverse labour market outcomes. Although we do not detect any effect on the decision to have children, we find that female students exposed to high-achieving peers change their fertility timing by having their first child at a younger age.

The remainder of the paper is organized as follows. Section 3.2 gives an overview of the related literature. Section 3.3 describes the institutional settings at CBS during our period of interest. Section 3.4 describes the data and our analysis sample. Section 3.5 describes our empirical strategy. Section 3.6 shows our results. Last, section 3.7 concludes.

3.2 Related Literature

Our study relates to two main research areas. We address the extensive literature on gender gaps in the labour market by providing evidence that the social environment in higher education matters for differences in career outcomes of male and female students. Particularly, as many students in our sample join the top of the Danish income distribution, we contribute to the literature on gender gaps among top earners (Bertrand, 2018, Boschini et al., 2020, Fortin et al., 2017).

Experimental studies have investigated competitiveness as one of the many explanations behind the gender gap. These papers find that females tend to shy away from competition (e.g., Niederle and Vesterlund (2007), Vandegrift and Yavas (2009) and Datta Gupta et al. (2013)). Similarly, women tend to avoid competitive negotiations (Babcock & Laschever, 2009) and apply less for competitive jobs (Flory et al., 2015). When exposed to competitive environments, females do not increase their performance as men do (Gneezy et al., 2003). Leadership, top level and managerial positions are often associated with competitiveness (see, for example,Niederle and Vesterlund (2007)). This paper do not speak directly to any mechanism. However, the finding that higher ability peers hurts female outcomes can be interpreted in the light of this literature on competitiveness. We contribute to this literature by showing that a competitive environment, as indicated by high-achieving peers, in a non-experimental setting in a higher education study program has long-lasting effects through students careers.

Second, we contribute to the large literature on the role of peers for the returns to educational programs (e.g., Chetty et al., 2011, Abdulkadiroglu et al., 2014). Recent studies have shown that peer ability matters for educational outcomes, and that responses are heterogeneous with respect to gender. Fischer (2017) investigates a quasi-random assignment process of first-year college students to chemistry classes and finds that females paired with higher ability peers are less likely to graduate in a STEM field. Using within-school across-cohort variation, Cools et al., 2019 proxy peer ability by their mothers' education and report that a higher share of high achieving male peers in high school hurts female students in a range of academic outcomes, lowers their labour market participation and increases fertility. Mouganie and Wang, 2020 utilize the same identification strategy to study peer effects in STEM enrollment and report adverse enrollment effects of higher ability male peers on females, but positive effects of high-achieving females. We contribute to this growing literature on ability peer effects in education by focusing on gender differences in labour market outcomes.

Our paper is also related to the literature that shows that peer gender composition matters. Several studies find that exposure to female-dominated environments moves females towards more gender-stereotypical choices. For example, Zölitz and Feld, 2018 find that females in business school sections with higher shares of females choose less male-populated university majors. Brenøe and Zölitz, 2020 find that a higher share of females in Danish high school cohorts decreases female STEM enrollment, having lasting effects on occupations and wages. Last, Anelli and Peri, 2017 document a positive effect of high male shares in high schools in Milan on the probability to choose a male-dominated college major for male students, but do not document any lasting effects on income and occupations. We cannot address the gender share composition due to the stratification of the randomization in our setting. However, our results suggest that for female outcomes, the ability composition of the male peers matters.

The peer effects literature exploits various types of explicit random assignments to peers of different abilities in educational contexts (such as dormitories, courses or cohorts) (Feld and Zölitz, 2017, Bietenbeck, 2020, Sacerdote, 2001, Carrell et al., 2009). However, only a few papers combine both explicit randomization and a focus on labour market outcomes. Ribas et al., 2020 utilize a cohort assignment cut-off in a large Brazilian university and find that female students who are assigned to a cohort with peers of lower ability are more likely to work in management positions. Closest to the setting in our paper are a study by Feld and Zölitz, 2018, who combine data on randomized peer groups at Maastricht University with a labour market survey for recent graduates, and a study by M. Skibsted and Bjerge, 2016 who study the same institutional setting as we do, and look at the impact on starting wages. M. Skibsted and Bjerge, 2016 find that peers have limited adverse consequences for the students' starting wages. Feld and Zölitz, 2018, on the other hand, find that female students are severely hurt by higher ability peers of the opposite gender in terms of early labour market outcomes. We use detailed administrative data following the students' careers, and especially emphasize career achievements at the top of the income distribution that materializes many years post matriculation.

3.3 Business Economics at CBS in 1986-2006

Copenhagen Business School (CBS) is a large public Business School in Denmark's capital, Copenhagen. In this paper, we focus on CBS's largest study program; a three-

year-long degree in Business Economics. The program has been offered since 1929 at CBS. In the beginning, the program was a two-year full-time study with a high practical focus. However, at the start of our sample period it was a well-established three-year-long program with the objective to provide students with a theoretical background for solving economic, managerial, legal and organizational problems in the field of business economics with some aspects of the national economy as well. It was still oriented towards the private sector.
In 1993, as part of the Bologna process in Europe, students who graduated from the Business Economics program were officially awarded a Bachelor of Science degree (B.Sc.). Throughout the majority of the period we study, a degree from the Business Economics program was equated with a Bachelor degree from the US. E.g. a degree in Business Economics could be used for admission into a Master program. In 1986, for example, 80-90% of the students who graduated from the business economics program continued to either a Master of Science in Economics and Business Administration or a Master of Science in Business Economics and Auditing; both of these programs were equivalent to a Master of Science offered at universities in Denmark.[2]
Like other study programs in Denmark, the Business Economics program was free of charge. Students were eligible for a government paid stipend. However, the generosity of the stipend varied over the period we study.
In the sample period, CBS enrolled around 600-700 students in the program each year. Applications and admission were handled by the centralized admission system together with all applications to higher education in Denmark.

The institutional features of the Business Economics program is well suited for studying peer effects. Most importantly, incoming students were not allowed to chose their peers, they where assigned to peer groups of around 35-40 students.
The peer groups were assigned before the start of the first semester and allocated based on the only information available to the CBS administration - the social security number. From the social security number three criteria can be generated: gender, age, and if the student is a Danish citizen.[3] The CBS administration aimed at balanced groups based on gender and foreign citizenship, while older students were assigned together in specific groups. The peer group assignment was therefore (conditionally) as good as random.

We use this information on initial peer group assignment. The peer groups stayed together for the entire length of the program and, importantly, it was almost impossible for students to change their peer group assignment. Exceptions required a valid cause (e.g., overlap with scheduled medical treatment). In addition, the composition of the peer groups might also change due to resource allocation - around one-third of students did not continue in their initial peer group to graduation. In case of a significant drop-out, peer groups were occasionally merged but this only happened after the first year of study.

[2] In 2007 CBS became one of Denmark's 8 university institutions.

[3] Non-Danes who apply from abroad were given a pseudo social security number in their application. This differs from the regular social security number in a recognisable way.

Throughout the study, we refer to students who were initially assigned to the same peer group as "group peers", while we refer to students from the same matriculation cohort as "cohort peers".

The Business Economics program primarily consisted of mandatory courses in the three main subjects; national economics, business economics, and academic tools, such as statistics. Teaching was mostly organized as classroom teaching within the peer groups - similar to the usual teaching style known from high school.[4]

The first semester was sixteen weeks long, of which the first two weeks were introduction weeks. These two weeks were used as an general introduction to CBS but also to become acquainted with the other students in the peer group. In general, CBS highly encouraged a good atmosphere within the peer groups, and the study guidelines in 1986, for example, mentions that "the group is your fixed point of reference throughout the study".

Within the peer groups, students were also encouraged to form smaller reading groups. The study regulation for Business Economics at CBS from 1986 continues "within most groups, reading groups are formed during the fall of the first year. Reading groups [...] in general consist of 3-5 other students with whom you solve assignments, discuss the syllabus, exchange notes, etc. [...]." Further, it continues, "in addition, studying, and collaborating on home assignments and the like, can be more enjoyable and beneficial when the relationship is good".

Teaching was standardized across the different groups, and students faced the same exams and the same requirements. Teachers were assigned without taking any peer composition into account.

The semester had 7 mandatory courses and each week had around 20 classroom hours. The number of weekly classroom hours declined a bit for the later semesters to around 15 hours. Students were expected to spend 40 hours a week in total on their studies.

The amount of in-class teaching was relative stable over the period we study. In 1986, it was a total of 470 teaching hours in the first academic year, and in 1995 is was 459 teaching hours.

Elective courses were a new teaching style that was implemented in the beginning of the 80s and only few electives were offered. In 1986 elective courses were offered for 150 of the teaching hours (around 25%) in the final part of the studies; starting from the 4th semester.

3.4 Data

We combine administrative data on peer groups from CBS with register data hosted by Statistics Denmark.

Through the CBS administration, we have access to official records for 21 matriculation cohorts in Business Economics who were enrolled between 1986 and 2006. The

[4]The first year was exclusively based on this classroom teaching, while some of the courses during the second and third year were combined into a format more associated with lectures.

data provided by CBS contains information on the students' matriculation and ex-matriculation dates, admission GPA, citizenship, gender, age, and most notably the information on the peer group assignment.
Throughout our analysis, we keep student observations regardless of what happens post matriculation - e.g. irrespective of graduation status - and we estimate intention-to-treat peer effects.

From the registers at Statistic Denmark we collect information on the students background characteristics; this includes their prior educational attainment and work experience, their current place of residence, and their parents educational and income status. We measure all background variables in the year prior to matriculation at CBS.
We also use the registers to define our outcome variables of interests. Our main variable of interest is the annual wage earned in wage-employment for the highest paid job in a given year. This is constructed using the Danish linked employer-employee data. We follow this information over the students' career post matriculation. This means that we capture the potential immediate wage benefit if a student chooses to drop out from the Business Economics program.
We also use information on labor market attachment, employment status, occupation (specifically, management positions using the DISCO classification), as well as information on employers and workplaces. We use the latter to construct measures using the entire population of Danish workers. E.g. we calculate an occupation's average daily wage, its average daily hours, and define occupations as being dominated by males or females.
Further, we construct measures irrespective of wage-employment status. This includes disposable income and top income status defined as being part of the top 1% or top 10% of the income distribution of all Danish taxpayers in a given year.
We construct variables to address educational outcomes based on the graduation status in the Business Economics program, and graduation status post matriculation for any program in the Danish educational system. Finally, we construct information on family investment based on marital status in a given year, and information on children.

3.4.1 Sample Selection and Peer Groups

We restrict our sample to the students' first spell in the Business Economics program. The sample includes 12716 students covering the cohorts from 1986-2006, and 360 randomized peer groups.[5] Our main variable of interest is the admission GPA that we use to construct the peer ability measure. Due to a data break in 1992, the CBS data does not contain this information. However, we can recover 74% of this missing information from the registers at Statistics Denmark (these are for students in academic high school track in Denmark).[6]

[5]For some years, CBS allocated students with a background in marketing in the same groups. These were not part of the randomization process and we have therefore excluded them from our sample.
[6]Our results are not sensitive to whether we exclude the cohort of 1992

Table 1 shows the sample information at the cohort level. The average group size is 36 students with a standard deviation of 6 students. The group size has changed over the years with a tendency that group sizes increased over time. The explanation for the standard deviation within cohort is that the randomization was conducted before the students started at CBS, and some students, despite accepting a program offer, never show up. These students are not part of our sample.

Table 1: Cohorts and Peer Groups

Cohort	Students in cohort	Share of female students	Share with no admission GPA	Number of peer groups	Students in peer group - average (sd)
1986	629	40%	5%	24	26 (2)
1987	569	43%	1%	23	25 (3)
1988	657	47%	5%	20	33 (3)
1989	635	45%	3%	20	32 (4)
1990	646	42%	2%	20	32 (3)
1991	615	32%	4%	20	31 (3)
1992	617	35%	26%	20	31 (1)
1993	734	32%	5%	20	37 (2)
1994	646	29%	3%	18	36 (3)
1995	608	30%	2%	17	36 (3)
1996	569	28%	1%	18	32 (3)
1997	521	27%	2%	14	37 (2)
1998	557	33%	2%	14	40 (2)
1999	553	34%	2%	14	40 (3)
2000	584	31%	2%	14	42 (2)
2001	604	36%	1%	14	43 (2)
2002	581	35%	< 1%	14	42 (3)
2003	581	31%	< 1%	14	42 (2)
2004	577	30%	1%	14	41 (2)
2005	622	28%	< 1%	14	44 (4)
2006	611	30%	< 1%	14	44 (3)
All	12716	34%	3%	360	36 (6)

3.4.2 Summary Statistics

Table 2 shows summary statistics for the pre-program information we have on the students in our sample. The table shows averages by male and female students.
The students in our sample are, on average, 21-22 years of age at matriculation and 95% are students with a Danish citizenship. Admission GPA is slightly higher for female students.
More than half of the sample took the academic high school track and another one-third took the business high school track. The students enter the Business Economics program, on average, 1.7 years after they graduated from high school. In the period between high school and the Business Economics program, 5% graduated from a higher education, and the students have accumulated around one year of work experience before entering the Business Economics program.
The students are from remarkably wealthy backgrounds; in the year prior to matricula-

Table 2: Summary Statistics: Background Variables

	Male	Female	Total
Age at matriculation	21.49	21.40	21.46
	(2.332)	(2.401)	(2.356)
Danish citizenship	0.955	0.951	0.953
	(0.208)	(0.216)	(0.211)
Admission GPA	8.364	8.466	8.399
	(0.803)	(0.722)	(0.778)
Academic high school track	0.568	0.593	0.576
	(0.495)	(0.491)	(0.494)
Business high school track	0.349	0.335	0.344
	(0.477)	(0.472)	(0.475)
Other high school track	0.0834	0.0724	0.0796
	(0.276)	(0.259)	(0.271)
Gap years prior to matriculation	1.728	1.711	1.722
	(2.001)	(1.981)	(1.994)
Higher degree prior to matriculation	0.0426	0.0579	0.0479
	(0.202)	(0.234)	(0.214)
Work experience prior to matriculation	1.027	0.957	1.003
	(1.136)	(1.118)	(1.130)
Father's income rank	86.07	86.50	86.22
	(20.56)	(20.12)	(20.41)
Mother's income rank	71.19	71.27	71.22
	(22.62)	(22.31)	(22.51)
Farther in top 10%	0.634	0.635	0.634
	(0.482)	(0.482)	(0.482)
Mother in top 10%	0.216	0.222	0.218
	(0.411)	(0.416)	(0.413)
Father's years of education	13.64	13.59	13.62
	(2.777)	(2.805)	(2.787)
Mother's years of education	12.98	12.65	12.87
	(2.679)	(2.741)	(2.705)
Father with a Master degree	0.176	0.165	0.172
	(0.381)	(0.371)	(0.378)
Mother with a Master degree	0.0711	0.0545	0.0654
	(0.257)	(0.227)	(0.247)
Father with a Bachelor degree	0.187	0.191	0.188
	(0.390)	(0.393)	(0.391)
Mother with a Bachelor degree	0.262	0.226	0.249
	(0.440)	(0.418)	(0.433)
Observations	8349	4367	12716

Mean summary statistics by male and female students. Standard deviations are reported in parenthesis.

tion the fathers of the students were on average ranked in the 86th income percentile within the Danish income distribution. The mothers of our students were ranked 71th. 63% of fathers and 22% of mothers had an income in the top 10% of the distribution. The parents had, on average, 13-14 years of education at the time our students started the Business economics program.

Figure 1 shows the GPA distribution for males and females. We see that the distribution for female students stochastically dominates the distribution for male students except for at the top (above approximately 1.5 scaled GPA).

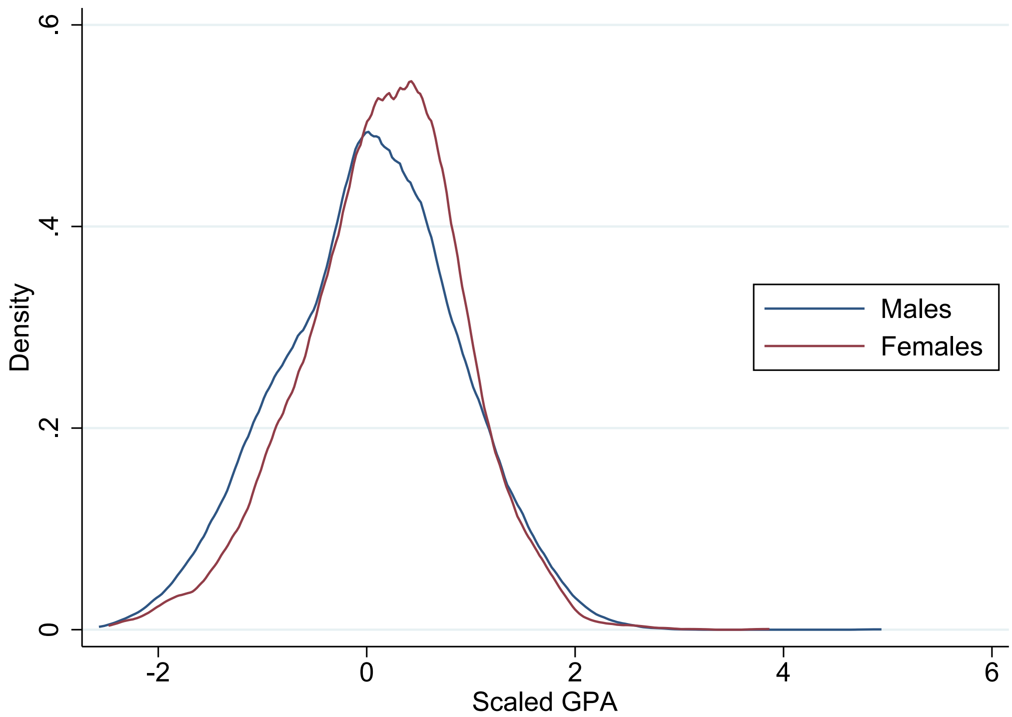


Figure 1: Standardized GPA distribution

Notes: The figure shows the standardized GPA distribution by gender.

Figure 2 sheds light on the differences in labor market outcomes for our male and female sample post matriculation. Figure 2 [a] shows that average annual wages are higher for the male sample compared to the female sample, and the increase in gender differences in earnings over time is particularly evident in the top of the distribution. The 90th percentile of the male sample strikingly outgrows the corresponding percentile of the female sample, and the mean wage of the male sample exceeds the 90th percentile of the female sample 10 years post matriculation.
Figure 2 [b] shows the differences in income by illustrating the share of the students that are top earners in a given year post matriculation, as defined by being in the top 10 or top 1 of the Danish income distribution. While 35% of female students reach the 90th percentile 10 years after matriculation, around 50% of the male students pass this

threshold within 10 years of matriculation. The gap between male and female students persists over the 25 years post matriculation.
Last, figure 2 [c] illustrates the share of students in management positions post-matriculation. The male students are more likely to work as managers, a difference that increases over time. For top-management positions, the same pattern is observed.

Table 3: Summary Statistics: Outcome Variables

	(1) Male	(2) Female	(3) Total
In wage employment	0.810	0.813	0.814
	(0.392)	(0.390)	(0.389)
Annual wage	419510.9	325265.3	385169.8
	(456779.9)	(287576.2)	(406987.8)
Hourly wage	322.6	251.9	298.2
	(1750.0)	(232.8)	(1399.5)
Disposable income	472554.7	373394.2	437031.1
	(930191.4)	(1108625.4)	(998919.1)
Labor market experience	7.413	7.665	7.451
	(6.092)	(6.221)	(6.146)
Management position	0.107	0.0602	0.0906
	(0.310)	(0.238)	(0.287)
Top management position	0.0233	0.00803	0.0180
	(0.151)	(0.0893)	(0.133)
Top 10% income distribution	0.382	0.289	0.349
	(0.486)	(0.453)	(0.477)
Top 1% income distribution	0.0678	0.0191	0.0503
	(0.251)	(0.137)	(0.219)
Observations	170213	94734	268346

Table 3 shows summary statistics for the post-matriculation outcomes for the students in our sample; Column (1) reports separately for male students, (2) separately for female student, and (3) combined.

Table 3 shows summary statistics for the pooled years post matriculation. 81% are in wage employment in a given year. The male sample earnings are almost 100.000 DKK more a year compared to the female sample. Large differences are also observed in disposable income and hourly wages. These statistics are despite that the male and female sample have similar accumulated work experience.

The students work in managerial positions in 9% of the yearly observations, and in around 2% of the yearly observations they work in top management positions. Again, we observe large gender differences. Our male sample work as managers in around 11% of the observations and for the female sample it is 6%. For top-management positions, the shares are 2.3% for the male sample, and only 0.8% for the female sample. A similar pattern is observed for top income shares. The careers in our sample represent the high end of the income distribution in Denmark - 35% of the disposable income observations in our panel are in the top 10% of the Danish income distribution, and 5% are in top 1%. However, females are 9.3 percentage points less likely to be observed in the top 10% and 3.5 times less likely to be observed in the top 1%.

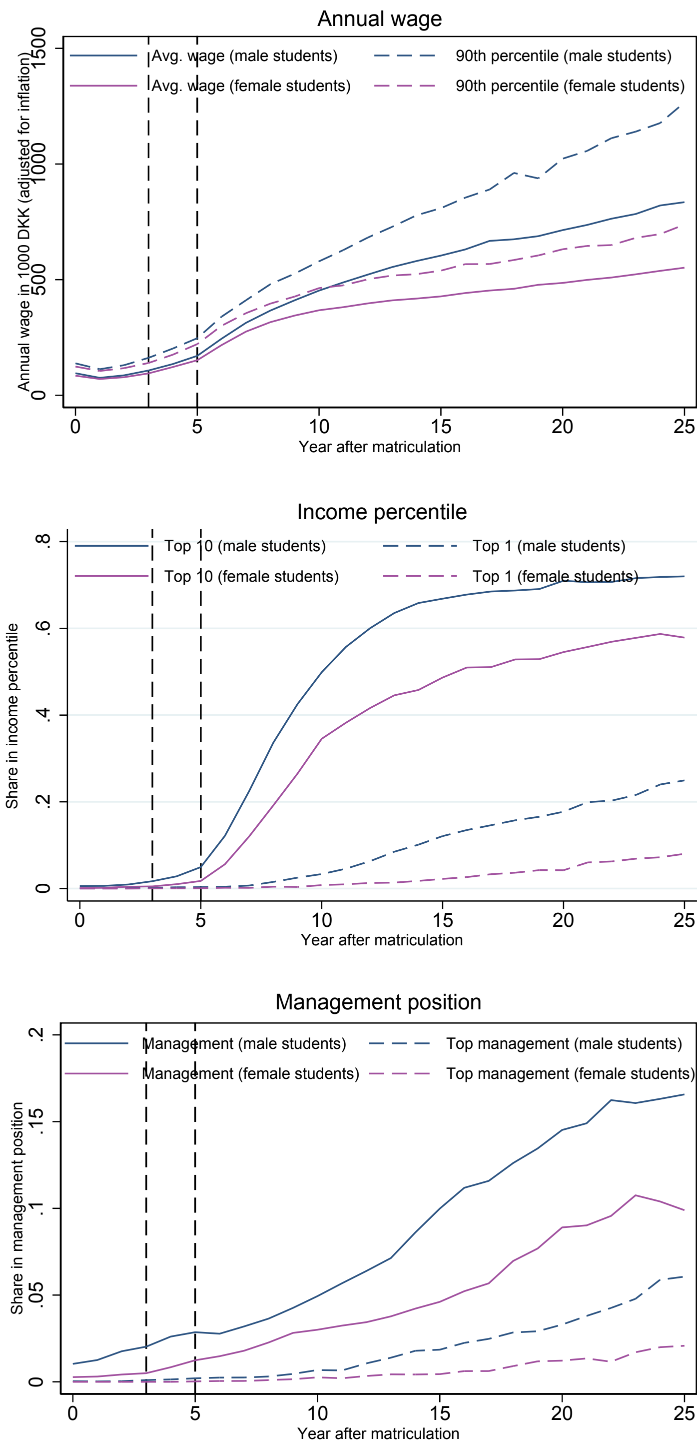


Figure 2: Outcome Variables by Gender Post Matriculation

Notes: The figure shows the following statistics for the sample of students by years after matriculation: In the top figure is the mean and 90th percentile annual wage, in the middle figure is the share of students at the top 1% and top 10% of the income distribution, and in the bottom figure is the share of students in management and top management positions.

3.4.3 Admission GPA

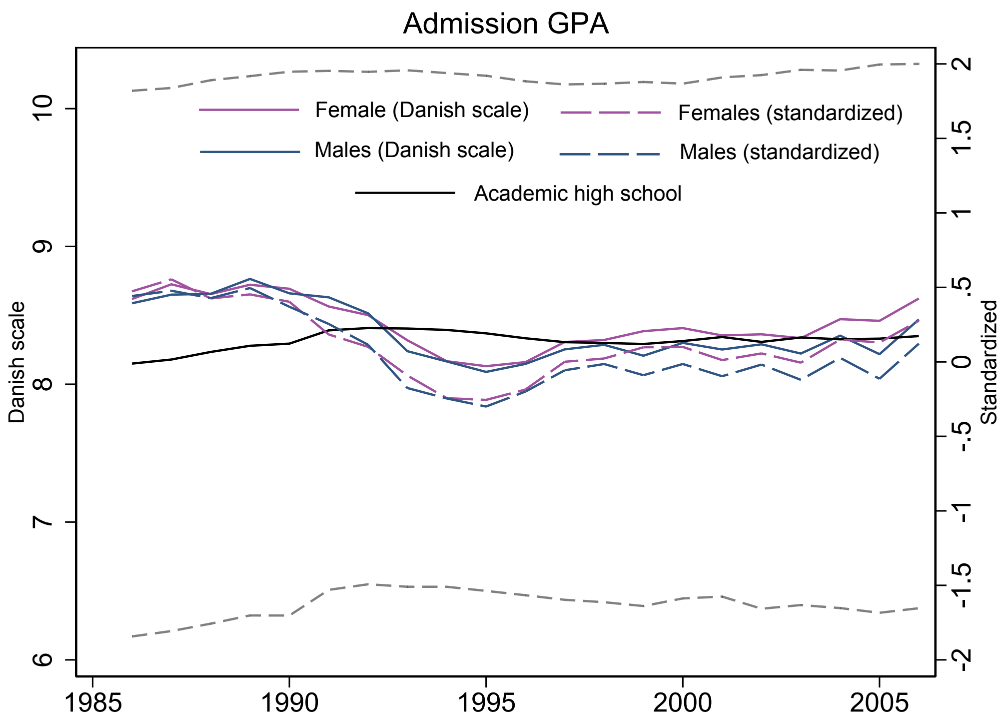


Figure 3: Admission GPA

Notes: The figure shows the mean of admission GPA across cohorts. The right-scale shows the mean using the Danish grading scale, and the left-scale the mean standardized by the academic high school cohort. The black line shows the high school mean for reference with the dotted lines indicating the +/- 2 standard deviations from the mean

We measure peer ability using the pre-determined GPA that the students used in the admission process. At the time of our sample period, the grading scale in Denmark had 10 steps with a numerical range from 0-13. Figure 3 shows the mean of the admission GPA over the cohorts using the Danish grading scale on the left axis. We standardize admission GPA using the distribution of all (academic) high school students in Denmark corresponding to the matriculation year at CBS[7]. Figure 3 shows in solid (dashed) black the mean (+/- 2 standard deviation) of the Danish academic high school students. The average high school GPA is relative stable around 8.1-8.3. In the right axis, the figure shows the mean of the standardized GPA for the CBS students. Around 1993 CBS increased the number of available spots in Business Economics and the average GPA declined for the subsequent cohorts. Before this period the average was in the high end of the Danish high school GPA distribution, and the average has again increased for the

[7]We only use the distribution for academic high school tracks due to lack of information from other high school tracks in the beginning of our sample period

later cohort to around the Danish average.

3.5 Empirical Strategy

The identification of peer effects is impeded by a set of well-known challenges (Manski (1993)). First, in real-world settings, students often choose their peer groups, which leads to a problem of self-selection. Second, outcomes within peer groups might be subject to common shocks. Last, the reflection problem emerges since peers mutually affect each others' outcomes. A typical solution for the reflection and the common shocks problem is to use predetermined peer characteristics as a treatment, and we use a predetermined peer ability measure. To address concerns regarding endogenous sorting into peer groups based on unobservables, a broad set of studies have examined students that are randomly assigned to their peers (Sacerdote (2001), Feld and Zölitz (2017)). In our context, CBS randomizes students to their initial peer groups within cohorts. Therefore, conditional on a set of stratification variables, the initial composition of peer groups is considered as good as random.

3.5.1 Empirical Specification

Our analysis relies on the natural random variation in peer group ability across the assigned peer groups within a given cohort. Figure 4 shows the variation that we use. It is standard in the peer effects literature to model peer ability by the leave-one-out

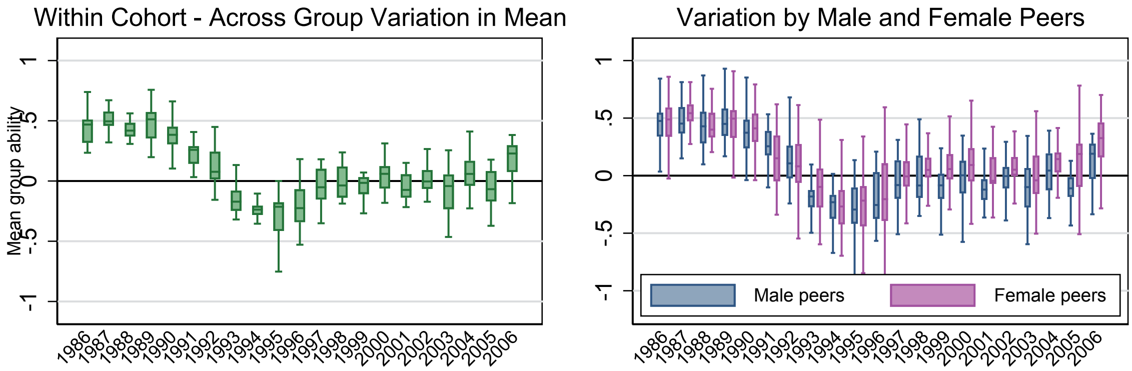


Figure 4: Identifying Variation

Notes: The figure shows the identifying variation. In the first figure is the variation that we use for our main specification. It shows the mean ability variation within cohort, across the randomized groups. The second figure shows the same variation by male and female peers.

mean:

$$\overline{GPA}_{-i,g} = \frac{1}{N_g - 1} \sum_{i}^{N_g} GPA_{-i}, \tag{3.1}$$

where N_g is the number of peers in student i's peer group g and GPA_i is student i's admission GPA.
It is also standard to include students own ability measure in the model to avoid bias from the mechanical negative correlation between own ability and the construction of peer ability. This is known as exclusion bias (Caeyers & Fafchamps, 2016).

We are interested in the possible differential impact of peer ability on male and female students, and for our main specification we use the following linear-in-means regression model[8]:

$$\begin{aligned} y_{i,gc,t} &= \alpha_{1:m} Male_i \times \overline{GPA}_{-i,g} + \alpha_{2:m} Male_i \times GPA_i \\ &+ \alpha_{1:f} Female_i \times \overline{GPA}_{-i,g} + \alpha_{2:f} Female_i \times GPA_i \\ &+ \gamma' X_i + \lambda_c + \psi_t + \varepsilon_{i,gc,t}, \end{aligned} \tag{3.2}$$

where $y_{i,gc,t}$ is the outcome variable of interest for student i in peer group g within cohort c in year t after matriculation. We focus on intention-to-treat estimation and all years post matriculation are included in the pooled regression. All regressions include controls for the stratification covariates; age, gender, and an indicator for Danish citizenship, in X_i. We include cohort fixed-effects, λ_c, and year after matriculation fixed-effects, ψ_t. All standard errors are clustered at the level of the randomization; cohorts.[9]

The main parameters of interests are $\alpha_{1:m}$ and $\alpha_{1:f}$ of the leave-one-out mean peer ability. They are identified from the (conditional) random variation into peer groups. We test the validity of the randomization in the next section.

3.5.2 Test of random assignment

The identification of our parameters of interest crucially depends on the variation in peer GPA to be uncorrelated with other observed (and unobserved) determinants of students' careers. The literature has suggested several tests to check if peer groups are indeed assigned "at random" (see, for example, Caeyers and Fafchamps (2016) for an overview). Usually, such tests try to demonstrate that a given predetermined leave-one-out peer group characteristic - here $\overline{GPA}_{-i,g}$ - is uncorrelated with a set of the students

[8]Given the focus of our study is on the long term aspect of assignment to peer ability, our main specification is the standard leave one-out-mean. We will return to other moments of the ability distribution in section 3.6.4.

[9]Due to the small number of clusters (G = 21), we report both cluster robust standard errors, and wild cluster bootstrapped p-values for the key parameters of interest (Cameron et al., 2008, Roodman et al., 2019).

pre-determined covariates; $X_{i:pre}$. To provide evidence in favour of this exogeneity, we follow three approaches for balancing and sorting tests.

The first approach - the balancing test - checks directly if the variation in peer GPA is correlated with any observed predetermined characteristics of the students. The test takes the same form as in Eq.3.2, but substitutes outcomes y_{ict} with individual-level variables that are determined prior to matriculation. The intuition of the test is that if peer groups are randomized they should not predict pre-determined characteristics since any correlation should be broken by the randomization.

$$\begin{aligned} X_{i:pre} &= \alpha^b_{1:m} Male_i \times \overline{GPA}_{-i,g} + \alpha^b_{2:m} Male_i \times GPA_i \\ &+ \alpha^b_{1:f} Female_i \times \overline{GPA}_{-i,g} + \alpha^b_{2:f} Female_i \times GPA_i \\ &+ \gamma' X_i + \lambda_c + \varepsilon_{i,gc}, \end{aligned} \tag{3.3}$$

Table 4 describes the results of the test. While - as expected - student's own high school GPA is correlated with many of the student's background characteristics, all of the variables except one appear to be (conditionally) uncorrelated with the treatment variable; male-peer GPA or female-peer GPA. The one variable which is marginally significant on the 5% level is Father in Top 10%. We see that females with high achieving peers are slightly more likely to have a father in the top 10% of the danish income distribution. With the large number of variables we test, there is a good chance that we would find one significant variable by chance.

The second approach tests if similar students are sorted into the same peer groups. The absence of significant correlations between peer and individual characteristics is then interpreted as a sign of randomness in the group assignment process. A set of simple regression of the leave-one-out peer characteristic on the pre-determined individual covariates is commonly referred to as a "naive" test, as such a test ignores a potential mechanical negative correlation between the peer-group variable, *Peer* GPA_i, and the student's individual GPA (further explained below). We start with the naive approach and check for conditional correlation between students' own predetermined variable values and a leave-one-out average value for peer GPA.

According to Table 5, the "naive" test detects no sorting (neither positive nor negative). Still, the coefficients might be negatively biased, and the null hypothesis of the absence of positive sorting might be mistakenly declined. As noted by Guryan et al. (2009), even under random assignment into peer groups, it will be the case that the student's individual GPA's are negatively correlated with his/her group's leave-one-out statistic *Peer* GPA_i. This phenomenon occurs as students cannot be assigned to themselves - if a student with an extreme realization, e.g. very high GPA, is assigned to a peer group, the pool of remaining students that might be assigned as his peer peers will have lower GPA's. This negative correlation results in a negative bias, which is known as exclusion bias.

Table 4: Balancing Test

	Male × *Peer GPA*	*Female* × *Peer GPA*	*Male* × *GPA*	*Female* × *GPA*
High School	-0.000	-0.020	-0.012	0.016
	(0.992)	(0.689)	(0.255)	(0.311)
Business High School	-0.011	0.020	-0.005	-0.031*
	(0.761)	(0.700)	(0.677)	(0.073)
Other HS equivalent	0.011	-0.000	0.016***	0.015***
	(0.468)	(0.991)	(0.000)	(0.000)
Gap Years	0.002	0.088	0.005	-0.046
	(0.990)	(0.512)	(0.909)	(0.365)
Post-HS degree	-0.028	-0.010	0.010***	0.001
	(0.191)	(0.652)	(0.003)	(0.870)
Work Experience	-0.032	-0.028	-0.035***	-0.038***
	(0.242)	(0.382)	(0.000)	(0.000)
Captial region	-0.023	-0.004	-0.043***	-0.035***
	(0.483)	(0.925)	(0.000)	(0.002)
Father's Income Rank	-1.113	-0.987	1.412***	2.348***
	(0.487)	(0.616)	(0.001)	(0.000)
Mother's Income rank	-0.805	-2.346	0.797*	0.849
	(0.574)	(0.332)	(0.053)	(0.210)
Father in Top 10%	-0.045	-0.075	0.019***	0.051***
	(0.301)	(0.115)	(0.007)	(0.002)
Father in Top 1%	0.029	0.065**	0.017***	0.014*
	(0.320)	(0.044)	(0.002)	(0.063)
Mother in Top 10%	0.017	0.045	0.015**	0.021**
	(0.383)	(0.102)	(0.048)	(0.032)
Mother in Top 1%	-0.000	0.001	0.003	-0.001
	(0.991)	(0.948)	(0.229)	(0.649)
Father's Education	-0.023	-0.143	0.104*	0.104
	(0.908)	(0.562)	(0.059)	(0.261)
Mother's Education	0.180	-0.004	0.180***	0.192***
	(0.355)	(0.986)	(0.004)	(0.003)

Each row in the table represents a balance test using the specification in equation 3.3, e.g. the first row uses a dummy for a post-high school degree as the (predetermined) dependent variable. The columns report the estimated coefficients for α^b's from equation 3.3. P-values in parenthesis are calculated with the wild cluster bootstrap (5000 replications) on the cohort level. All specifications include cohort fixed effects, and the stratification variables; gender, Danish citizenship, and age. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Sorting Test

	”Naive” Test	Split-Sample Test
High School	0.003	0.004
	(0.285)	(0.945)
Business High School	0.003	0.003
	(0.228)	(0.957)
Other HS equivalent	0.001	0.002
	(0.640)	(0.984)
Gap Years	-0.007	-0.009
	(0.266)	(0.488)
Post-HS degree	0.001	0.001
	(0.788)	(0.994)
Work Experience	0.005	0.006
	(0.104)	(0.939)
Captial region	0.004	0.003
	(0.168)	(0.946)
Father’s Income Rank	-0.004	-0.003
	(0.154)	(.)
Mother’s Income Rank	0.001	0.001
	(0.773)	(.)
Father in Top 10%	0.002	0.002
	(0.444)	(0.969)
Father in Top 1%	-0.003	-0.003
	(0.338)	(0.960)
Mother in Top 10%	-0.001	-0.001
	(0.577)	(0.987)
Mother in Top 1%	0.003	0.004
	(0.281)	(0.983)
Father’s Education	0.004	0.005
	(0.270)	(0.509)
Mother’s Education	0.001	0.002
	(0.871)	(0.819)

The first column of the table reports results of the ”naive” sorting test - the coefficients from a regression of leave-one-out mean of a peers’ background variable on the corresponding own background variable. The second column represents coefficients and p-values calculated using the split-sample method developed by Stevenson (2015). P-values in parenthesis are calculated with the wild cluster bootstrap on the matriculation cohort level. All specifications include cohort fixed effects, and the stratification variables; gender, Danish citizenship, and age. It was not possible to get p-values for the split-sample estimator for the variables Father’s Income Rank and Mother’s Income Rank due to an issue which can arise when peer groups are large (Stevenson, 2015). *** p<0.01, ** p<0.05, * p<0.1.

The third approach implements a split-sample randomization test as suggested in Stevenson (2015) to handle the exclusion bias.[10] The idea of the test is to break the mechanical correlation by splitting the sample. Stevenson (2015) recomends randomly sampling one student from each peer group to use in the estimation, but notes that as peer groups grow in size it becomes more difficult to estimate the variance of the split sample estimator. To get around the difficulties with estimating the variance one can include more students from each peer group, but this does however introduce some bias. Stevenson (2015) finds that the introduced bias is small in her simulations. As our peer groups are large in size we implement the procedure in the following way. First, we pick five students per group randomly, then calculate group averages using the other students and finally calculate conditional correlations. The resulting coefficients should be less mechanically biased, but less precisely estimated due to the reduced size of the estimation sample. Therefore we repeat the same procedure 5,000 times and aggregate the estimates using the formulas provided by Stevenson (2015). As we see from the second column of the Table 5, correcting for exclusion bias does not change the conclusion from the naive test, although we are not able to estimate variances for the split sample estimator for father's and mother's income rank.[11] Even though the coefficients from the split-sample procedure tend to be higher, none of them are significant at the 10% level. We therefore conclude that students do not appear to be sorted into groups.

3.6 Results

This section starts with our main results on key labour market outcomes in section 3.6.1; the section looks at labor market earnings, labor supply, and the effect of high-achieving peers on achieving "top positions" - such as earning incomes in the top 1% and 10% of the income distribution or working in a position with management responsibilities.
In section 3.6.2 we decompose our peer measure into the mean peer ability of male and female students, respectively, and in section 3.6.3 we look more into heterogeneity by the students own pre-determined ability measure. In section3.6.4 we examine other parts of the peer ability distribution in addition to the mean.
Last, in section 3.6.5, we shed light on two potential reasons for our negative findings on female careers; we explore educational outcome and family investment.

3.6.1 Effects on Labour Market Outcomes

Table 6 presents results for the causal effect of leave-one-out mean peer ability on log annual earnings. Column (1) presents our preferred specification that includes controls for the stratification variables; *Danish citizenship, gender and age*, the students

[10] Stevenson (2015) shows that the properties of her suggested split-sample test dominate the more common exclusion bias correction in Guryan et al. (2009).

[11] We have tried to increase the number of randomly selected students and to increase the number of repetitions, but have not been able to solve the problem.

Table 6: Labour Market Earnings with Specification Checks

	(1) Annual earnings	(2) Annual earnings	(3) Annual earnings	(4) Annual earnings
Male$\times$ $\overline{GPA}_{-i,g}$ ($\hat{\alpha}_{1:m}$)	-0.020	-0.001	-0.022	-0.023
	(0.030)	(0.037)	(0.030)	(0.040)
Female$\times$ $\overline{GPA}_{-i,g}$ ($\hat{\alpha}_{1:f}$)	-0.094**	-0.089**	-0.096**	-0.085*
	(0.035)	(0.037)	(0.035)	(0.044)
Male$\times$ GPA_i ($\hat{\alpha}_{2:m}$)	0.070***	0.068***	0.070***	0.069***
	(0.005)	(0.005)	(0.005)	(0.005)
Female$\times$ GPA_i ($\hat{\alpha}_{2:f}$)	0.038***	0.044***	0.038***	0.037***
	(0.010)	(0.012)	(0.010)	(0.010)
Specification				
Stratification controls	✓	✓	✓	✓
Cohort + year after matriculation FE	✓	✓	✓	✓
Individual controls		✓		
Peer controls			✓	
Gender interaction with FE				✓
Mean dependent variable				
Male students	12.445	12.445	12.445	12.445
Female students	12.285	12.285	12.285	12.285
Bootstrapped p-value for H_0:				
$\hat{\alpha}_{1:m} = 0$	0.527	0.974	0.485	0.585
$\hat{\alpha}_{1:f} = 0$	0.018	0.023	0.016	0.071
$\hat{\alpha}_{1:m} = \hat{\alpha}_{1:f}$	0.057	0.054	0.060	0.335
Observations	212,683	177,346	212,683	212,683
R-squared	0.503	0.514	0.503	0.504

The dependent variable in all columns is the natural logarithm of annual earnings based on the highest paid job in the given year. Annual earning are adjusted by inflation. Stratification controls include age, gender and Danish citizenship. Furthermore, all specifications includes cohort and year-after-matriculation fixed effects. Standard errors clustered at the cohort level are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1. Column (2) adds individual controls including all the background variables reported in Table 2. Column (3) adds peer controls including gender share and size of peer group. Column (4) interacts gender with the fixed-effects. The second to last panel in the table shows wild cluster bootstrapped p-values for the effect of peer ability on males and females under the null hypothesis of no effects and a test for whether those they are statistically different from each other.

own admission GPA interacted with gender, as well as the set of cohort and year-after-matriculation fixed effects. As the identifying variation is as good as random conditional on the stratification controls within cohort our estimates should not be sensitive to the inclusion of additional controls. We have included the other specifications to illustrate how our estimates are affected by including additional controls. In columns (2) - (4), we gradually add more controls to the model. Column (2) adds additional individual controls and column (3) peer group controls. Column (4) includes gender interactions with the fixed effects. Adding controls does not affect the sife of the estimated parameters much, which again confirms the random assignment. The coefficients are relatively stable, and a one standard deviation increase in mean peer ability reduces female students earnings by 9-10%. The estimate for male students are close to zero and insignificant in all specifications. Note, that one standard deviation increase is large relative to the natural variation that we have from the randomization (see Figure 4).
In addition, the GPA controls show that male students see larger returns on their own admission GPA compared to female students.

Table 7 shows the effect of peer ability on other outcomes of interest in the labour market. All five columns are based on our preferred specification, column (1) in table 6. Our measure of annual wage earnings is conditional on wage employment. Column (1) looks at peer effects on the probability of wage employment in a given year. Around 81% of the students in our sample are observed in wage employment and we find a negative peer effect for female students: a one standard deviation higher peer ability decreases the probability of wage employment by 3.5 percentage points or a 4.3% reduction compared to the baseline level.[12] We find no effect for male students. Column (2) shows no effect on the number of days employed, conditional on being wage-employed.
Column (3) replicates the finding in column (1) in Table 6, and in column (4) we only look at annual wages after the students have graduated from their highest observed degree. Column (5) shows the result for disposable income. This outcome measure is not conditional on wage employment. The results on disposable income mirror the results on earnings and with a sizeable increase in magnitude. An increase in peers' average ability by one standard deviation decreases female income by almost 20%. Note that the effect on income potentially reflects various margins of adjustment, including the margin of employment, self-employment and non-labour income. Last, column (6) shows the effect on hourly wages. Our results are not sensitive to how and when we measure wages.

Table 8 reports the effects of having high-achieveing peers on what we call top outcomes for male and female students. Column (1) and (2) show the impact of peer ability on the probability of being in a management or top management position, and column (3) and (4) show the impact on being in the top 10% or top 1% of the Danish income distribution. The estimates in column (1) and (2) say that a one standard deviation increase in mean peer ability decreases the likelihood to work as a manager by 4.1 percentage points for female students, e.g. 68.3% compared to baseline, and by 1.8

[12] Out of wage-employment is a combined state of non-employment, self-employment, unemployment or in education

Table 7: Labour Market Outcomes

	(1) Wage employed	(2) Days employed	(3) Annual earnings	(4) Annual earnings*	(5) Income	(6) Hourly wage
Male× $\overline{GPA}_{-i,g}$ ($\hat{\alpha}_{1:m}$)	-0.013	-4.053	-0.020	-0.022	-0.050	0.005
	(0.011)	(2.700)	(0.030)	(0.032)	(0.034)	(0.022)
Female× $\overline{GPA}_{-i,g}$ ($\hat{\alpha}_{1:f}$)	-0.035***	-3.411	-0.094**	-0.085**	-0.190***	-0.062***
	(0.011)	(2.519)	(0.035)	(0.038)	(0.031)	(0.019)
Male× GPA_i ($\hat{\alpha}_{2:m}$)	0.002	1.888***	0.070***	0.077***	0.072***	0.074***
	(0.002)	(0.537)	(0.005)	(0.005)	(0.006)	(0.004)
Female× GPA_i ($\hat{\alpha}_{2:f}$)	0.005*	0.590	0.038***	0.043***	0.038***	0.049***
	(0.003)	(0.836)	(0.010)	(0.010)	(0.010)	(0.004)
Specification						
Stratification Controls	✓	✓	✓	✓	✓	✓
Cohort and year after matriculation FE	✓	✓	✓	✓	✓	✓
Mean dependent variables						
Male students	0.813	311.728	12.445	12.509	12.609	5.611
Female students	0.815	310.284	12.285	12.354	12.522	5.466
Bootstrapped p-value for H_0:						
$\hat{\alpha}_{1:m} = 0$	0.245	0.159	0.521	0.502	0.173	0.796
$\hat{\alpha}_{1:f} = 0$	0.005	0.177	0.021	0.033	0.000	0.004
$\hat{\alpha}_{1:m} = \hat{\alpha}_{1:f}$	0.038	0.814	0.062	0.085	0.002	0.006
Observations	260,780	207,255	212,683	181,048	234,500	164,210
R-squared	0.103	0.071	0.503	0.503	0.350	0.354

The dependent variables in the order of the columns are; an indicator for being wage-employed in a given year, number of days in employments, the natural logarithm of annual earnings, the natural logarithm of annual earnings after graduation from highest degree, the natural logarithm of annula income, and the natural logarithm of hourly wage. All specifications includes cohort and year-after-matriculation fixed effects as well as the stratification controls age, gender and Danish citizenship. Standard errors clustered at the cohort level are in parenthesis. The table also reports means of the dependent variables by gender. The second to last panel shows wild cluster bootstrapped p-values for the effect of mean peer ability on males ($\hat{\alpha}_{1:m}$) and mean peer ability on females ($\hat{\alpha}_{1:f}$) under the null hypothesis of no effects, and a test of whether those two effects are statistically different under the null that $\hat{\alpha}_{1:m} = \hat{\alpha}_{1:f}$. *** p<0.01, ** p<0.05, * p<0.1.

percentage points for the top management positions, e.g. 225% compared to baseline.[13] In column (3) and (4) we show the effect on being in the top of the income distribution among all Danish tax payers. Females with higher ability peers are less likely to reach the top of the income distribution: a one standard deviation increase in peer ability decreases female students' prospects of entering the 90th income percentile by 4.6 percentage points, e.g. 15.9% compared to baseline. Similarly, female students are 3.9 percentage points less likely to join the top 1% of earners in Denmark, which translates into 205.3% compared to the baseline estimate of 1.9%.

Figure 5 illustrates the effect of high-achieving peers on male and female by year post matriculation. We see that the adverse effects of peers on female careers appear after some years in the labour market.

[13] Note that the linear probability model might provide a misleading intuition of the magnitude. A future version will include estimates from Logit models as well.

Table 8: Top Outcomes

	(1) Manager	(2) Top Manager	(3) Top 10%	(4) Top 1%
$Male \times \overline{GPA}_{-i,g}$ $(\hat{\alpha}_{1:m})$	0.011	0.008	-0.016	0.007
	(0.012)	(0.007)	(0.014)	(0.010)
$Female \times \overline{GPA}_{-i,g}$ $(\hat{\alpha}_{1:f})$	-0.041***	-0.018**	-0.046***	-0.039***
	(0.012)	(0.008)	(0.015)	(0.009)
$Male \times GPA_i$ $(\hat{\alpha}_{2:m})$	0.007**	0.004***	0.054***	0.026***
	(0.002)	(0.001)	(0.004)	(0.002)
$Female \times GPA_i$ $(\hat{\alpha}_{2:f})$	0.005	0.003*	0.050***	0.006***
	(0.003)	(0.002)	(0.005)	(0.002)
Specification				
Stratification Controls	✓	✓	✓	✓
Cohort and year-after-matriculation FE	✓	✓	✓	✓
Mean dependent variables				
Male students	0.108	0.024	0.382	0.068
Female students	0.060	0.008	0.289	0.019
Bootstrapped p-value for H_0:				
$\hat{\alpha}_{1:m} = 0$	0.374	0.260	0.274	0.492
$\hat{\alpha}_{1:f} = 0$	0.006	0.052	0.008	0.000
$\hat{\alpha}_{1:m} = \hat{\alpha}_{1:f}$	0.003	0.008	0.047	0.001
Observations	170,649	170,649	236,055	236,055
R-squared	0.068	0.035	0.340	0.095

The dependent variables in the order of the columns are; an indicator for being in a management position, in a top management position, the natural logarithm of disposable income in a given year, an indicator of being in the top 10% of the Danish income distribution in a given year, and an indication of being in the top 1% of the Danish income distribution in a given year. Stratification controls include age, gender and Danish citizenship. Furthermore, all specifications includes cohort and year-after-matriculation fixed effects. Last, the table shows wild cluster bootstrapped p-values for the effect of peer ability under the null hypothesis of no effects, and a test if those two effects are statistically different from each other. Standard errors clustered at the cohort level are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

3.6.2 Gender Decomposed Peer Ability:

Two recent papers, Feld and Zölitz (2018) and Cools et al. (2019), suggests that the gender-composition of the peer ability distribution matters for outcomes of male and female students. In this section, we decompose our measure of peer ability into the average ability of a student's male peers and average ability of a student's female peers and run the specification:

$$\begin{aligned} y_{i,gc,t} &= \alpha^m_{1:m} Male_i \times \overline{GPA}^m_{-i,g} + \alpha^f_{1:m} Male_i \times \overline{GPA}^f_{-i,g} + \alpha_{2:m} Male_i \times GPA_i \\ &+ \alpha^m_{1:f} Female_i \times \overline{GPA}^m_{-i,g} + \alpha^f_{1:f} Female_i \times \overline{GPA}^f_{-i,g} + \alpha_{2:f} Female_i \times GPA_i \\ &+ \gamma' X_i + \lambda_c + \psi_t + \varepsilon_{i,gc,t}, \end{aligned} \tag{3.4}$$

where $\overline{GPA}^m_{-i,g}$ is constructed as the leave-one-out mean of student i's male peers

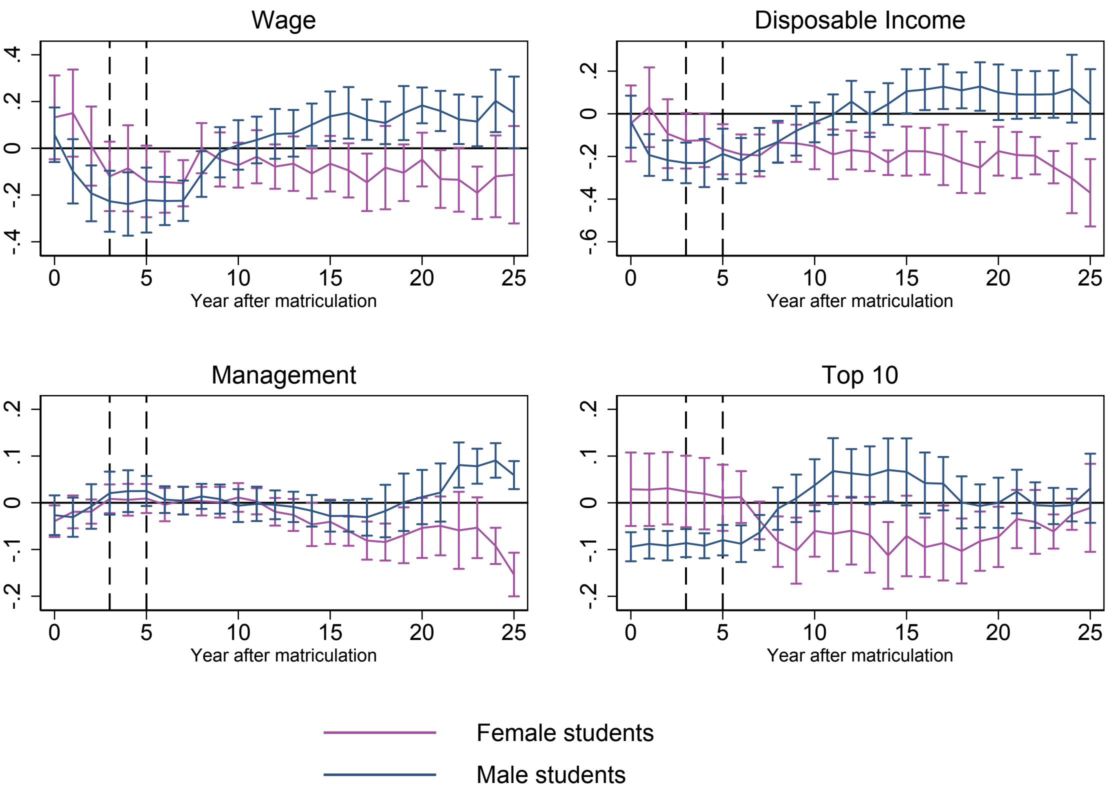


Figure 5: Peer ability effect by year post matriculation

Notes: The figure shows the coefficients from our preferred specification with the mean peer ability interacted with year after matriculation.

leaving student i out, and $\overline{GPA}^{f}_{-i,g}$ is the leave-one-out mean of student i's female peers leaving student i out.

Overall we find that high-achieving male peers drive the negative effects for female students. Table 9 summarizes the results on a range of our outcome measures. In the first column we see that a one standard deviation increase in male peer-ability decreases males likelihood of being wage employed by 1.8% points and females likelihood by 2.1% points, whereas exposure to high-achieving female peers has no effect on neither females nor males. The size of the effect for males is close to what we estimated in table 7, although it was not significant there. Given all the other results an explanation could be that having high achieving male peers causes males to be more likely to enter into self-employment. In the second column we find that a one standard deviation increase in male peer-ability decreases female wages by 7.9%, whereas exposure to high-achieving female peers has no effect on neither females nor males. These results are in alignment with findings in Feld and Zölitz (2018) who also report severe negative effects for females of being assigned to higher ability males in terms of early labour market outcomes. In

Table 9: Gender-Decomposed Peer Ability

	(1) Wage employment	(2) Annual earnings	(3) Income	(4) Manager	(5) Top 10%
Male Peers:					
$Male \times \overline{GPA}^{m}_{-i,g}$ ($\hat{\alpha}^{m}_{1:m}$)	-0.018**	-0.008	-0.044	0.004	-0.009
	(0.008)	(0.025)	(0.030)	(0.010)	(0.014)
$Female \times \overline{GPA}^{m}_{-i,g}$ ($\hat{\alpha}^{m}_{1:f}$)	-0.021**	-0.079***	-0.144***	-0.036***	-0.020
	(0.009)	(0.026)	(0.021)	(0.010)	(0.013)
Female peers:					
$Male \times \overline{GPA}^{f}_{-i,g}$ ($\hat{\alpha}^{f}_{1:m}$)	0.008	-0.005	0.010	0.011	-0.006
	(0.007)	(0.019)	(0.016)	(0.007)	(0.011)
$Female \times \overline{GPA}^{f}_{-i,g}$ ($\hat{\alpha}^{f}_{1:f}$)	-0.012	-0.003	-0.029	0.001	-0.022**
	(0.009)	(0.027)	(0.023)	(0.009)	(0.010)
Specification					
Stratification Controls	✓	✓	✓	✓	✓
Cohort and year after matriculation FE	✓	✓	✓	✓	✓
Mean dependent variables					
Male students	0.813	12.445	12.609	0.108	0.382
Female students	0.815	12.285	12.522	0.060	0.289
Bootstrapped p-value for H_0:					
Male peers:					
$\hat{\alpha}^{m}_{1:m} = 0$	0.039	0.786	0.175	0.711	0.529
$\hat{\alpha}^{m}_{1:f} = 0$	0.025	0.011	0.000	0.002	0.131
Female peers:					
$\hat{\alpha}^{f}_{1:m} = 0$	0.277	0.808	0.554	0.167	0.600
$\hat{\alpha}^{f}_{1:f} = 0$	0.191	0.915	0.254	0.956	0.039
Observations	260,780	212,683	234,500	170,649	236,055
R-squared	0.103	0.503	0.350	0.068	0.340

The dependent variables in the order of the columns are; an indicator for being wage-employed in a given year, the natural logarithm of annual earning, the natural logarithm of disposable income, an indicator for being in a management position, and an indicator for being in the top 10% of the Danish income distribution. All specifications includes cohort and year-after-matriculation fixed effects and stratification controls age, gender and Danish citizenship. Standard errors clustered at the cohort level are in parenthesis. The table also reports the mean dependent variables by gender. Last, the table shows wild cluster bootstrapped p-values for the effect of mean peer ability on males ($\hat{\alpha}_{1:m}$) and mean peer ability on females ($\hat{\alpha}_{1:f}$) under the null hypothesis of no effects, and a test if those two effects are statistically different under the null that $\hat{\alpha}_{1:m} = \hat{\alpha}_{1:f}$. *** p<0.01, ** p<0.05, * p<0.1.

contrast to Feld and Zölitz (2018), we do not find any significant and positive effects of increasing the average ability of male peers on males for labour market outcomes.
The pattern is consistent for the other outcomes measuring disposable income, and management positions. The exception is the effect on being in top 10 of the Danish income distribution. The effect on female students of an increase in either male or female peer ability are in magnitude associated with a 2.2 percentage point decline. However, it is only the part driven by female peers that is significant on the 5% level.

3.6.3 Heterogeneity by student's own pre-determined ability measure

Studies have shown that in a competitive environment it is often females in the high end of the ability distribution that are most negatively affected. In this section, we examine if the estimated peer effects on male and female students are heterogeneous with respect to the student's own ability.
We investigate if higher ability peers differentially affect students with different pre-determined ability. To do so, we interact our peer effect specification with an indicator if a student is above or below the median of the distribution of admission GPAs' within his or her cohort. The results of this exercise can be found in Table 10.

Separating low and high ability males does not produce any new insights. For both groups of males, we do not find evidence in support of the hypothesis that peer ability matters for outcomes except for a positive effect on being in a management position for male students in the lower end of the ability distribution and negative effects on earnings and the likelihood of full time employment for males in the upper end of the ability distribution, although both estimates are only significant on the 10% level.
For female students, the previous results are reinforced. For a range of outcome variables, the effect of increasing peer-ability is sizeable and negative for female students of both high and low ability. Further we find that the adverse effects for high-ability females are more pronounced: they experience a wage penalty of -14.1% for a one standard deviation increase in peer ability. This is significantly higher than the wage decrease for low ability females of -5.1% (insignificant). For the other outcome variables, such as reaching the Top 10% of the income distribution, or working as a manager, the effect on high-ability females is larger as well. The exception is the effect on being in wage employment, where the negative effect is larger in magnitude for the lower end of the ability distribution.

3.6.4 Other peer ability moments

In our previous specifications, we have assumed that peer effects operate through the linear-in-mean ability specification. Given that the literature suggests other non-linear peer effects in the educational contexts (e.g. Anelli and Peri (2017), Mouganie and

Table 10: Peer Effects by Students Own Ability Type

	(1) Wage employment	(2) Annual earnings	(3) Income	(4) Manager	(5) Top 10%
Student GPA below median within cohort:					
$Male \times \overline{GPA}_{-i,g}$ $(\hat{\alpha}^{below}_{1:m})$	-0.021	0.023	-0.039	0.033*	-0.001
	(0.013)	(0.038)	(0.038)	(0.016)	(0.017)
$Female \times \overline{GPA}_{-i,g}$ $(\hat{\alpha}^{below}_{1:f})$	-0.042***	-0.051	-0.163***	-0.034**	-0.015
	(0.014)	(0.045)	(0.037)	(0.014)	(0.019)
Student GPA above median within cohort:					
$Male \times \overline{GPA}_{-i,g}$ $(\hat{\alpha}^{above}_{1:m})$	-0.005	-0.064*	-0.064	-0.011	-0.034*
	(0.013)	(0.034)	(0.043)	(0.015)	(0.017)
$Female \times \overline{GPA}_{-i,g}$ $(\hat{\alpha}^{above}_{1:f})$	-0.028**	-0.141***	-0.216***	-0.048***	-0.073***
	(0.013)	(0.041)	(0.033)	(0.015)	(0.019)
Specification					
Stratification Controls	✓	✓	✓	✓	✓
Cohort and year after matriculation FE	✓	✓	✓	✓	✓
Mean dependent variables					
Female students below median	0.811	12.270	12.514	0.058	0.259
Male students below median	0.812	12.393	12.566	0.103	0.342
Female students above median	0.824	12.298	12.534	0.064	0.315
Male students above median	0.822	12.489	12.658	0.116	0.422
Bootstrapped p-value for H_0:					
$\hat{\alpha}^{below}_{1:m} = 0$	0.137	0.560	0.332	0.056	0.957
$\hat{\alpha}^{below}_{1:f} = 0$	0.007	0.272	0.002	0.026	0.465
$\hat{\alpha}^{above}_{1:m} = 0$	0.735	0.089	0.171	0.463	0.070
$\hat{\alpha}^{above}_{1:f} = 0$	0.044	0.006	0.000	0.008	0.003
$\hat{\alpha}^{below}_{1:m} = \hat{\alpha}^{above}_{1:m}$	0.309	0.126	0.358	0.113	0.176
$\hat{\alpha}^{below}_{1:f} = \hat{\alpha}^{above}_{1:f}$	0.031	0.017	0.000	0.020	0.007
Observations	260,780	212,683	234,500	170,649	236,055
R-squared	0.103	0.503	0.350	0.068	0.340

The dependent variables in the order of the columns are; an indicator for being wage-employed in a given year, the natural logarithm of annual earning, the natural logarithm of disposable income, an indicator of being in a management position, and an income of being in top 10% of the Danish income distribution. All specifications includes cohort and year-after-matriculation fixed effects and the stratification controls age, gender and Danish citizenship. Standard errors clustered at the cohort level are in parenthesis. The table also reports the means for the dependent variables by gender and whether a student is below or above the median GPA. The second to last panel shows wild cluster bootstrapped p-values for the effect of mean peer ability on males ($\hat{\alpha}_{1:m}$) and mean peer ability on females ($\hat{\alpha}_{1:f}$) under the null hypothesis of no effects, and a test if those two effects are statistically different under the null that $\hat{\alpha}_{1:m} = \hat{\alpha}_{1:f}$. *** p<0.01, ** p<0.05, * p<0.1.

Wang (2020)), we go beyond the linear peer effects model in this section.

We use ability shares instead of a parametric distribution. We follow Mouganie and Wang (2020) and classify students within the cohort in three groups; we classify "high-ability" as students with an admission GPA in the top 20% of the cohort distribution, "low-ability" as students in the bottom 20 of the cohort distribution, and "median-ability" as student in the range in between. Based on this classification, we construct peer-ability measures using the leave-on-out proportion within each classification.

Table 11: Non-Linear Effects

	(1) Wage employ-ment	(2) Annual earning	(3) Income	(4) Manager	(5) Top 10%
Male × *Share Low Ability* ($\hat{\alpha}^{low}_{1:m}$)	0.007	0.145	0.022	0.027	0.046
	(0.068)	(0.088)	(0.031)	(0.036)	(0.035)
Female × *Share Low Ability* ($\hat{\alpha}^{low}_{1:f}$)	-0.047	0.109	0.019	0.059	0.019
	(0.120)	(0.101)	(0.036)	(0.043)	(0.046)
Male × *Share High Ability* ($\hat{\alpha}^{high}_{1:m}$)	-0.038	0.008	-0.022	0.030	0.009
	(0.081)	(0.098)	(0.020)	(0.036)	(0.038)
Female × *Share High Ability* ($\hat{\alpha}^{high}_{1:f}$)	-0.155*	-0.165	-0.026	0.022	-0.052
	(0.086)	(0.109)	(0.034)	(0.037)	(0.045)
Specification					
Stratification Controls	✓	✓	✓	✓	✓
Cohort and year after matriculation FE	✓	✓	✓	✓	✓
Mean dependent variables					
Female students	12.285	12.525	0.818	0.061	0.290
Male students	12.443	12.614	0.817	0.109	0.383
Bootstrap p-value H_0:					
$\hat{\alpha}^{low}_{1:m} = 0$	0.919	0.125	0.497	0.466	0.210
$\hat{\alpha}^{low}_{1:f} = 0$	0.700	0.298	0.596	0.204	0.685
$\hat{\alpha}^{high}_{1:m} = 0$	0.642	0.937	0.300	0.429	0.801
$\hat{\alpha}^{high}_{1:f} = 0$	0.110	0.163	0.445	0.593	0.280
$\hat{\alpha}^{low}_{1:f} = \hat{\alpha}^{low}_{1:m}$	0.931	0.147	0.615	0.309	0.372
$\hat{\alpha}^{high}_{1:f} = \hat{\alpha}^{high}_{1:m}$	0.271	0.401	0.272	0.625	0.497
Observations	212,683	234,500	260,784	171,159	236,055
R-squared	0.502	0.350	0.106	0.066	0.340

Notes: The dependent variables in the order of the columns are; an indicator for being wage-employed in a given year, the natural logarithm of annual earning, the natural logarithm of disposable income, an indicator of being in a management position, and an income of being in top 10% of the Danish income distribution. All specifications includes cohort and year-after-matriculation fixed effects and the stratification controls age, gender and Danish citizenship. Standard errors clustered at the cohort level are in parenthesis. The table also reports means for the dependent variables by gender. The second to last panel shows wild cluster bootstrap p-values for the effect of peer ability on males and females under the null hypothesis of no effects and a test if those two effects are statistically different from each other. Standard errors clustered at the cohort level are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

The results for this exercise are depicted in Table 11. We find no evidence supporting the hypothesis that the negative labour market returns for females are driven by a higher share of high-ability peers in their respective peer group for any of the dependent variables.

3.6.5 Educational and Family Investment

In the previous sections, we have shown that high-achieving peers have a severely negative impact on female students' careers. Several life investments that potentially explain these adverse effects come to mind. First, female students who are assigned to higher ability peers might do worse in the labour market because those same peers affect their *human capital accumulation*; higher ability peers might harm female performance and their graduation rates (Feld and Zölitz (2018), Mouganie and Wang (2020)), leading to fewer women entering top jobs (Frederiksen & Kato, 2018).

Second, females who are exposed to high achieving peers might decide to prioritize having a family over a career if exposure to many higher ability male peers reinforces traditional gender roles (Brenøe and Zölitz, 2020, Cools et al., 2019). Exposure to male peers of higher ability might also reduce search frictions in partnership formation and accelerate family investments for females (Skyt Nielsen & Svarer, 2009).
Fertility in itself is an interesting outcome measure for two reasons: first, if high achieving peers lead females to shy away from entering the contest for top jobs, a higher rate of fertility might be a natural byproduct as many of these jobs are less compatible with having children. Otherwise, peers might alter females' preferences towards motherhood, and the negative effects of peers on labour market outcomes might be explained by female students having (more) children (Lundborg et al. (2017), Kleven et al. (2019)). Previous studies in different educational contexts suggest that classroom composition might have effects on female fertility: Cools et al. (2019) find that more male peers of higher ability increase female fertility at the intensive margin at age 26-32, whereas more high achieving female peers decrease male's fertility. Brenøe and Zölitz (2020) reports that a higher share of female peers in high school changes females' timing of the first child as well as their number of children.

Table 12 investigates how peer ability is related to educational outcomes. Conditional on graduating from the program, female students who are exposed to peers of higher ability finish the Business Economics program with lower GPA: an increase of peer admission GPA by one standard deviation decreases female students final study program GPA by around 11.8% of a standard deviation.
We do not find any significant effects og having high achieving peers on graduating from the business economics program, although the estimated coefficient is large (-10.9%) and corresponds in sign and magnitude to findings in M. K. Skibsted (2016), who investigates the effect of peer ability on graduation rates in the same study program for a subset of our sample.
For graduating with any bachelor degree we do not find any effects on female students.

Table 12: Education

	(1) University GPA	(2) Graduate	(3) Bachelor	(4) Master	(5) Male Dominated
Male$\times \overline{GPA}_{-i,g}$ $(\hat{\alpha}_{1:m})$	-0.022	0.052	-0.053*	0.001	0.018
	(0.028)	(0.075)	(0.030)	(0.034)	(0.026)
Female$\times \overline{GPA}_{-i,g}$ $(\hat{\alpha}_{1:m})$	-0.118***	-0.109	0.005	-0.111***	-0.107***
	(0.032)	(0.089)	(0.026)	(0.036)	(0.029)
Male$\times GPA_i$ $(\hat{\alpha}_{2:m})$	0.085***	0.664***	-0.013**	0.098***	0.020
	(0.009)	(0.017)	(0.005)	(0.010)	(0.013)
Female$\times GPA_i$ $(\hat{\alpha}_{2:f})$	0.076***	0.601***	-0.037***	0.097***	0.044***
	(0.012)	(0.030)	(0.010)	(0.011)	(0.011)
Specification					
Stratification Controls	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓
Mean dependent variables					
Female students	0.649	0.017	0.267	0.517	0.466
Male students	0.679	-0.009	0.238	0.501	0.524
Bootstrap p-value for H_0:					
$\hat{\alpha}_{1:m} = 0$	0.443	0.501	0.086	0.966	0.501
$\hat{\alpha}_{1:f} = 0$	0.001	0.246	0.857	0.008	0.002
$\hat{\alpha}_{1:f} = \hat{\alpha}_{1:m}$	0.005	0.123	0.069	0.001	0.002
Observations	12,293	7,597	10,967	10,967	10,021
R-squared	0.063	0.239	0.033	0.125	0.407

Notes: The dependent variables in the order of the columns are; the standardized final university GPA, an indicator for whether student graduates from the Business Economics Program at CBS, an indicator for whether the student ever graduates with any bachelor degree, and indicator for whether the student ever graduates with any master degree, and an indicator for whether more than 60% of students in the subsequent educational program are males. All specifications include cohort fixed effects and the stratification controls age, gender, and Danish citizenship. Further we control for own GPA in all specifications. The table also reports the mean for all dependent variables by gender. The second to last panel shows wild cluster bootstrap p-values for the effect of peer ability on males and females under the null hypothesis of no effects and a test if those two effects are statistically different from each other. Standard errors clustered at the cohort level are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

While we find a negative effect on male students, although it is only significant on the 10% level. We see a large negative effect for female students for graduating with any master degree: an increase of peer admission GPA by one standard deviation decreases female students likelihood of graduating with any master degree by 11.1% points. Last, female students exposed to high achieving peers are also less likely to pursue a male dominated subsequent education degree.

Table 13 presents results for the effect of peer abilitty on marriage and fertility outcomes. In columns one and two, we first examine the peer ability effect on marriage and being in a partnership. For both male and female students we observe a negative effect of being assigned to highers achieving peers. For female students the effect is only significant for the extended definition that includes living in a partnership.
Table 13 also examines fertility patterns. In column (3) we find that male students exposed to high achieving peers are more likely to have children in a given year, but find

Table 13: Family Investment

	(1) Married	(2) Married*	(3) Children	(4) Child at age 30	(5) Age at First Child
$Male \times \overline{GPA}_{-i,g}$ ($\hat{\alpha}_{1:m}$)	-0.044**	-0.048**	0.031**	-0.013	0.001
	(0.020)	(0.017)	(0.013)	(0.032)	(0.372)
$Female \times \overline{GPA}_{-i,g}$ ($\hat{\alpha}_{1:f}$)	-0.032	-0.039**	0.012	0.024	-0.829**
	(0.020)	(0.017)	(0.017)	(0.029)	(0.351)
$Male \times GPA_i$ ($\hat{\alpha}_{2:m}$)	0.003	0.008**	0.002	-0.001	-0.100*
	(0.003)	(0.004)	(0.003)	(0.004)	(0.051)
$Female \times GPA_i$ ($\hat{\alpha}_{2:m}$)	-0.000	-0.002	0.010*	-0.021**	0.121
	(0.007)	(0.007)	(0.005)	(0.009)	(0.115)
Specification					
Stratification Controls	✓	✓	✓	✓	✓
Cohort FE	✓	✓	✓	✓	✓
Year-after-matriculation FE	✓	✓	✓		
Mean dependent variables					
Male students	0.311	0.555	0.624	0.218	32.014
Female students	0.381	0.596	0.544	0.325	30.921
Bootstrapped p-value for H_0:					
$\hat{\alpha}_{1:m} = 0$	0.033	0.007	0.025	0.678	0.998
$\hat{\alpha}_{1:f} = 0$	0.131	0.038	0.484	0.425	0.032
$\hat{\alpha}_{1:m} = \hat{\alpha}_{1:f}$	0.462	0.609	0.070	0.114	0.025
Observations	257,490	257,490	257,490	12,293	8,890
R-squared	0.305	0.082	0.459	0.020	0.044

The dependent variables in the order of the columns are; an indicator for being married in a given year, an indicator for being married or in a partnership in a given year, an indicator of having children in a given year, an indicator for having a child by age 30, and the students age when they got their first child conditional on having a child. Stratification controls include age, gender and Danish citizenship. All specifications includes cohort fixed effects, and columns (1)-(3) also include year-after-matriculation FE. Further, all specifications control for own GPA. The second to last panel shows wild cluster bootstrapped p-values for the effect of peer ability under the null hypothesis of no effects, and a test if those two effects are statistically different from each other. Standard errors clustered at the cohort level are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

no effect for female students. Further, we find no effects on having a child by age 30 for either male of female students. By conditioning on having a child, column (5) of Table 13 shows that exposure to peers of higher ability reduces a female student's age at first childbirth, suggesting that they shift their fertility decisions. While we cannot provide any conclusive evidence, family investments does not seem to play any major role in the negative consequence for female students.

3.7 Conclusion

In this paper, we make use of the random assignment of students to peer groups in a large Danish Business Economics program - covering matriculation cohorts from 1986 to 2006 - to investigate how exposure to higher ability peers contributes to the gender divergence in the top earning distribution.
We show that females exposed to high-achieving peers tend to be less successful in their

careers in terms of lower earnings and weaker labour market attachment. The negative effect is also evident in a lower probability to reach leading positions in firms (management or top management positions), and the top of the income distribution (top 1% or top 10%).

Furthermore, we find that females of higher ability are particularly hurt by exposure to higher ability peers and by decomposing our measure of peer ability into parts associated with male and female peers, we find that higher ability male peers mostly drive the adverse effects on females.

To shed light on potential reasons for the negative effect, we investigate educational outcomes and family investment. Female students who are exposed to peers of higher ability are significantly less likely to graduate from the Business Economics program, and are less likely to obtain any higher education degree. Conditional on graduation, they have lower GPA than their cohort peers.
We find limited evidence that the assignment to peers of higher ability leads female students to change family investments.

Given this, we conclude that our results provide evidence that the composition of university peers affects labour market outcomes with long lasting consequence. Although we are not able to fully disentangle the underlying behavioural reasons, our evidence is, however, consistent with the experimental literature showing that female students are less competitive. This might affect female students' when exposed to higher achieving peers. So before thinking about policy implications of our results, we believe that it is essential to gain a better understanding of the behavioral aspect for the negative impact. We hope that documenting the long-lasting consequences will foster more research in the area.

References

Abdulkadiroglu, A., Angrist, J., & Pathak, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, *82*(1), 137–196.

Albrecht, J., Björklund, A., & Vroman, S. (2003). Is There a Glass Ceiling in Sweden? *Journal of Labor Economics*, *21*(1), 145–177.

Anelli, M., & Peri, G. (2017). The Effects of High School Peers' Gender on College Major, College Performance and Income. *The Economic Journal*, *129*(618), 553–602.

Atkinson, A. B., Casarico, A., & Voitchovsky, S. (2018). Top Incomes and the Gender Divide. *The Journal of Economic Inequality*, *16*(2), 225–256.

Babcock, L., & Laschever, S. (2009). *Women Don't Ask: Negotiation and the Gender Divide*. Princeton University Press.

Bertrand, M. (2018). Coase Lecture – The Glass Ceiling. *Economica*, *85*(338), 205–231.

Bietenbeck, J. (2020). The Long-Term Impacts of Low-Achieving Childhood Peers: Evidence from Project STAR. *Journal of the European Economic Association*, *18*(1), 392–426.

Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, *55*(3), 789–865.

Boschini, A., Gunnarsson, K., & Roine, J. (2020). Women in Top Incomes: Evidence from Sweden 1974–2013. *Journal of Public Economics*.

Brenøe, A., & Zölitz, U. (2020). Exposure to More Female Peers Widens the Gender Gap in STEM Participation. *Journal of Labor Economics, 38*(4).

Caeyers, B., & Fafchamps, M. (2016). *Exclusion Bias in the Estimation of Peer Effects* (tech. rep.). National Bureau of Economic Research.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based Improvements for Inference with Clustered Errors. *The Review of Economics and Statistics*, *90*(3), 414–427.

Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, *27*(3), 439–464.

Catalyst. (2020). *Pyramid: Women in S&P 500 Companies* (tech. rep.).

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, *126*(4), 1593–1660.

Cools, A., Fernández, R., & Patacchini, E. (2019). *Girls, boys, and high achievers* (Working Paper No. 25763). National Bureau of Economic Research.

Datta Gupta, N., Oaxaca, R. L., & Smith, N. (2006). Swimming Upstream, Floating Downstream: Comparing Women's Relative Wage Progress in the United States and Denmark. *ILR Review*, *59*(2), 243–266.

Datta Gupta, N., Poulsen, A., & Villeval, M. C. (2013). Gender Matching and Competitiveness: Experimental Evidence. *Economic Inquiry*, *51*(1), 816–835.

Feld, J., & Zölitz, U. (2017). Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. *Journal of Labor Economics*, *35*(2), 387–428.

Feld, J., & Zölitz, U. (2018). *Peers from venus and mars – higher-achieving men foster gender gaps in major choice and labor market outcomes* (Working Paper). In: Cesifo Area Conferences : Economics of Education.

Fischer, S. (2017). The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence. *Labour Economics*, *46*, 211–226.

Flory, J., Leibbrandt, A., & List, J. (2015). Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions. *Review of Economic Studies*, *82*(1), 122–155.

Fortin, N. M., Bell, B., & Böhm, M. (2017). Top Earnings Inequality and the Gender Pay Gap: Canada, Sweden, and the United Kingdom. *Labour Economics*, *47*, 107–123.

Frederiksen, A., & Kato, T. (2018). Human Capital and Career Success: Evidence from Linked Employer-Employee Data. *The Economic Journal*, *128*(613), 1952–1982.

Gallen, Y., Lesner, R. V., & Vejlin, R. (2019). The Labor Market Gender Gap in Denmark: Sorting Out the Past 30 Years. *Labour Economics*, *56*, 58–67.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics*, *118*(3), 1049–1074.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091–1119.

Guryan, J., Kroft, K., & Notowidigdo, M. J. (2009). Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. *American Economic Journal: Applied Economics*, *1*(4), 34–68.

Kleven, H., Landais, C., & Søgaard, J. E. (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, *11*(4), 181–209.

Lundborg, P., Plug, E., & Rasmussen, A. W. (2017). Can Women Have Children and a Career? IV Evidence from IVF Treatments. *American Economic Review*, *107*(6), 1611–37.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, *60*(3), 531–542.

Mouganie, P., & Wang, Y. (2020). High Performing Peers and Female STEM Choices in School. *Journal of Labor Economics*, *38*(3), 805–841.

Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

Olivetti, C., & Petrongolo, B. (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics*, *8*, 405–434.

Petrongolo, B., & Ronchi, M. (2020). Gender Gaps and the Structure of Local Labor Markets. *Labour Economics*, *64*.

Ribas, R. P., Sampaio, B., & Trevisan, G. (2020). Short- and Long-term Effects of Class Assignment: Evidence from a Flagship University in Brazil. *Labour Economics*, *64*.

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in stata using boottest. *The Stata Journal*, *19*(1), 4–60.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, *116*(2), 681–704.

Skibsted, M., & Bjerge, B. (2016). *Do Peers Matter ? - Impacts of Peers on Master ' s Choice and Labor Market Outcomes ∗*.

Skibsted, M. K. (2016). *Empirical essays in economics of education and labor*. Frederiksberg: Copenhagen Business School (CBS).

Skyt Nielsen, H., & Svarer, M. (2009). Educational Homogamy How Much is Opportunities? *Journal of Human Resources*, *44*(4), 1066–1086.

Smith, N., Smith, V., & Verner, M. (2011). The Gender Pay Gap in Top Corporate Jobs in Denmark: Glass Ceilings, Sticky Floors or Both? *International Journal of Manpower*, *32*(2), 156–177.

Stevenson, M. (2015). Tests of Random Assignment to Peers in the Face of Mechanical Negative Correlation: An Evaluation of Four Techniques. *University of Pennsylvania, Mimeo*.

Vandegrift, D., & Yavas, A. (2009). Men, Women, and Competition: An Experimental Test of Behavior. *Journal of Economic Behavior & Organization*, *72*(1), 554–570.

Zölitz, U., & Feld, J. (2018). The Effect of Peer Gender on Major Choice in Business School. *University of Zurich, Department of Economics, Working Paper*, (270).

CHAPTER 4
Peers and Careers: Labour Market Effects of Alumni Networks

with Alexander Fischer, Andrei Gorshkov, and Jeanette Walldorf

Abstract

Are social connections formed among university peers important in shaping their future careers? We answer this question by using records from the long-lasting random assignment of Business Economics students to peer groups at Copenhagen Business School from 1986 to 2006, which we merge to detailed labour market information from Danish registers. We find that students, randomly assigned to the same peer group, tend to have more similar careers than students from the same cohort, but a different peer group: they tend to work in the same occupations and industries and are more likely to be hired by the same employer. The strongest "excess" similarities of group peers over cohort peers are observed at the most disaggregated level, the workplace. This effect is strong, persistent (although decreasing over time), characterized by homophily and pronounced the most for students from the wealthiest family backgrounds. By comparing the transitions of students to workplaces with incumbent group peers to workplaces with incumbent cohort peers, we find that students benefit from their alumni network by gaining access to more stable and higher-paying jobs.

4.1 Introduction

It is a widely held belief that "knowing the right people" can have a transformative effect on one's career. Joining a network of professionally successful people is often considered to be an integral part of the value of an education program. Business schools - whose programs are often characterized as stepping stones into high paying careers - typically emphasize the importance of alumni peer ties.[1] Nevertheless, despite the broad

[1]Promotional materials of world-leading business schools often explicitly mention career benefits from networking among students (e.g., University of Chicago Booth (2018)).

academic interest in educational peer effects and general recognition of their importance (see Sacerdote (2014) for a review), still very little is known about the role of peers in shaping career outcomes. Moreover, the interpretation of potential peer effects on careers is intrinsically ambiguous. Interactions of educational peers could be limited to their shared period of studies, but still manifest in persistent labour market effects. On the other hand, peer interactions might extend into their professional lives, where peers from university form an alumni network, which acts as a source of information on jobs and career opportunities.

In this paper, we ask if and how social interactions in higher education affect student's career choices and if the social ties from university persist into labour market interactions of former academic peers. Investigating the effects of social interactions is notorious for its many challenges (see, for example, Manski (1993) and Angrist (2014)). The ideal research design would need to fulfil several restrictive requirements. Most importantly, in order to avoid ex-post similarities between a pair of students reflecting only sorting on ex-ante unobserved similarities, the social ties between students needed to be randomly assigned and independent of other factors that affect future careers. A second hurdle for investigating the effects of social interactions in higher education on student's careers is that data on labour market outcomes of students needs to be available. Given that it is difficult to find a setting that simultaneously satisfies both conditions, researchers often face a trade-off between the credibility of randomization and observability of detailed career dynamics.

The setting in our paper is characterized by both: arguably credible randomization and rich data on students' careers. First, we exploit a policy that randomly assigned students to peer groups in a Business Economics program at Copenhagen Business School (from now on abbreviated as CBS) that reaches as far back as 1986. Business Economics at CBS is a large business education program, and its graduates are typically among the country's top earners. Second, we can merge these records to Danish register data, which offer rich and detailed information on workers and firms in Denmark. Therefore, we can observe the individual career dynamics of students.

To identify the effect of peers on choices after graduation, we use a dyadic approach and compare pairs of students that are randomly assigned to the same peer group (group peers) to pairs of students from the same cohort but who are assigned to a different peer group (cohort peers). We find that group peers tend to have more similar careers than cohort peers. Group peers "excessively" tend to work in the same occupations and industries and are more likely to be hired by the same firm. The effects are strongest at the most disaggregated level - the workplace. During their first ten years after graduation, a pair of group peers is more than 40% more likely to be working in the same workplace than a pair of cohort peers. These workplace effects are much stronger than effects on occupation, industry and firm choice. Moreover, they seem to be the driving force underlying group peers' career similarities. Conditional on not working at the same workplace, group peers are not significantly more likely to hold jobs in the same firms or industries. Moreover, we find no evidence of peer similarities in educational choices that might mediate the effect, which is consistent with post-graduation network interactions between students. While our identification strategy addresses the selection problem, our empirical design does not rule out that results are driven by common group-level shocks. Nevertheless, we conclude that overall pattern of our findings suggest that the ”excess” career similarities between group peers (in comparison to cohort peers) is most likely indicative of social interactions with academic peers being impactful for students careers rather than just reflecting common shocks from the shared academic environment.

The social ties formed in university persist over time, but slowly fade out. Right after graduation, group peers are twice as likely to work together compared to cohort peers; ten years after scheduled graduation, the effect is attenuated to around 20%. This pattern is characterized by notable gender and country-of-origin homophily, with effects being more pronounced for peers with the same gender or country-of-origin and insignificant for ”dissimilar” peers. Moreover, the most substantial effects are observed for students with a higher initial ability (measured by high school GPA) and for students coming from the richest families (as measured by a paternal income in top 1% of the national income distribution in the year before matriculation).

This evidence suggests that students form active alumni networks after graduation and use them to gather information on job opportunities. Nevertheless, is joining a group peer's workplace also beneficial? To answer this question, we compare events where a student joins a workplace with one or more incumbent *peer group peers* to events where a student joins a workplace with one or more incumbent *cohort peers*. We find that such transitions are associated with getting higher wages and more stable jobs in better-paying firms, industries and occupations. In alignment with the results from the first part of the analysis, these returns are more pronounced at earlier career stages and for high-ability students from wealthy families.

This paper aims to contribute to the large and growing literature on peer effects in education. A prominent branch of research exploits experimental settings with various types of random assignments in educational context (within dormitories, courses or cohorts), but mostly focuses on academic outcomes (see, for example, work by Sacerdote (2001); D. Zimmerman (2003); Lyle (2007); Carrell et al. (2009) and Feld and Zölitz (2017)). A related paper by Bjerge and Skibsted (2016) studies peer effects on master's choice using the same data source and a similar approach as we do, although with a different sample. They find that peers are more likely to choose the same master's degrees.[2] A handful of studies attempt to establish a link between academic peers and long-run outcomes without explicit randomization by using across cohort variation (as in Hoxby (2000)). For example, Black et al. (2013) investigate how variation in socioeconomic composition of cohorts in ninth grade in Norway affects students' various long-run outcomes (including labour market outcomes). Anelli and Peri (2017) and Brenøe and Zölitz (2020) study the effects of class gender composition in Italian and Danish high schools, both with a focus on college major choices and income.

This project adds to a much smaller set of studies that combines both - explicit randomization of students into peer groups and a focus on labour market outcomes af-

[2]Our paper distinguishes it self from the paper by Bjerge and Skibsted (2016) as we are not focused on peer effects, but use higher intensity social interactions during the Business Economics degree to identify network effects in the labor market.

ter graduation. For example, using random assignment to sections in Harvard Business School, Lerner and Malmendier (2013) investigate how peers affect the decision to become an entrepreneur, and Shue (2013) explores the role of peer interaction between top managers in corporate decision making. Jones and Kofoed (2020) study how peers affect occupational preferences in the context of a military academy. The most related to the setting in our study is a paper by Feld and Zölitz (2018), which uses within-course randomization into classes in a Dutch business school to investigate the effects of peer GPA on course choices and evaluates labour market outcomes based on survey-based measures of career success and work satisfaction.

Our study's contribution to the literature on long-run peer effects is twofold. First, in contrast to the common interpretation of academic peer influences as being confined to the shared period of studies, we emphasize the role of contemporaneous alumni interactions in shaping students' careers by employing a labour market network perspective. Second, access to detailed administrative labour market data allows us to analyze in detail how academic peers affect each others' career choices at the industry, occupational, firm and workplace level.

This paper further contributes to a wide range of studies exploring the importance of social connections in the labour market. Who you know appears to be important in different social contexts: neighborhoods (Bayer et al. (2008); Hellerstein et al. (2011); Schmutte (2015)), former coworkers (Cingano and Rosolia (2012); Glitz (2017); Caldwell and Harmon (2018); Glitz and Vejlin (2019), Saygin et al. (2019)), family members (Kramarz & Skans, 2014) and ethnic groups (Edin et al. (2003), Damm (2009), Dustmann et al. (2016)). Several studies attempt to assess the importance of labour market networks among former students. Hacamo and Kleiner (2017) focus on the managerial market and investigate how firms use social connections that their employees gain through MBA programs to attract talent. Zhu (2018) identifies referral networks among graduates from community colleges in Arkansas.

Lastly, this study contributes by investigating a specific context in which peers are considered to be of particular importance for educational returns: business schools (e.g., Lerner and Malmendier (2013) and Shue (2013)). Besides identifying significant social interactions among former students and providing evidence of sizable benefits from such interactions, we point at significant disparities between students in the size of these effects. In accordance with (S. D. Zimmerman, 2019), we provide evidence that business education programs particularly benefit the accumulation of social capital among students from advantaged backgrounds. As a consequence, mobility into top incomes for initially disadvantaged students might be hindered by disproportionate network formation.

The rest of the paper is organized as follows. The next section discusses the institutional background of the Business Economics Program at CBS, data sources, and provides descriptive statistics. In Section 4.3, our strategy to identify peer effects as "excess similarities" is outlined, and the main results are presented. Section 4.4 investigates benefits from alumni interactions. The last section concludes.

4.2 Data and Institutional Background

4.2.1 Business Economics at CBS 1986-2006

Copenhagen Business School (CBS) is a large public Business School in Denmark's capital, Copenhagen. In this paper, we focus on CBS's largest study program; a three-year-long degree in Business Economics. The program has been offered since 1929 at CBS. In the beginning, the program it was a two-year full-time study with a high practical focus. However, at the start of our sample period it was a well-established three-year program with the objective to provide students with a theoretical background for solving economic, managerial, legal and organizational problems in the field of business economics with some aspects of the national economy as well. It was oriented towards the private section.

In 1993, as part of the Bologna process in Europe, students who graduated from the Business Economic program were officially awarded a Bachelor of Science degree (B.Sc.). Throughout the majority of the period we study, a degree from the Business Economics program was equated with a Bachelor degree from the US. E.g. a degree in Business Economic could be used for admission into a Master program. In 1986, for example, 80-90% of the students who graduated from the business economic program continued to either a Master of Science in Economics and Business Administration or a Master of Science in Business Economics and Auditing; both of these programs were equivalent to a Master of Science offered at universities in Denmark.[3]

Like other study programs in Denmark, the Business Economics program was free of charge. Students were eligible for a government paid stipend. However, the generosity of the stipend varied over the period we study.

In the sample period, CBS enrolled around 600-700 students in the program each year. Applications and admission were handled by the centralized admission system together with all applications to higher education in Denmark.

The institutional features of the Business Economics program is well suited for studying peer effects. Most importantly, incoming students were assigned to peer groups of around 35-40 students.

The peer groups were assigned before the start of the first semester and allocated based on the only information available to the CBS administration - the social security number. From the social security number three criteria can be generated: gender, age and if the student is a Danish citizen.[4] The CBS administration aimed at balanced groups based on gender and foreign citizenship, while older students were assigned together in specific groups. The peer group assignment was therefore (conditionally) as good as random.

We use this information on initial peer group assignment. These peer groups stayed

[3]In 2007 CBS became one of Denmark's 8 university institutions.

[4]Non-Danes who apply from abroad were given a pseudo social security number in their application. This differs from the regular social security number in a recognisable way.

together for the entire length of the program and, importantly, it was almost impossible to change the assigned group. The only few exceptions under which the composition of peer groups might change were either due resource allocation - around one-third of students do not continue to graduation. In case of a significant drop-out, peer groups were occasionally merged. Under other circumstances, it was almost impossible for students to change their peer group assignment. Exceptions require a valid cause (e.g., overlap with scheduled medical treatment).
Throughout the study, we refer to students who were initially assigned to the same peer group as "group peers", while we will refer to students from the same matriculation cohort as "cohort peers".

The Business Economics program primarily consisted of mandatory courses in the three main subjects; national economics, business economics, and academic tools, such as statistics. Teaching was mostly organized as classroom teaching within the peer groups - similar to the usual teaching style known from high school.[5]
The first semester was sixteen weeks long, of which the first two weeks were introduction weeks. These two weeks were used as an general introduction to CBS but also to get to know each other within the peer groups. In general, CBS highly encouraged a good atmosphere within the peer groups, and the study guidelines in 1986, for example, mentions that "the group is your fixed point of reference throughout the study".
Within the peer groups, students were also encouraged to form smaller reading groups. The study regulation for Business Economics at CBS from 1986 continues "within most groups, reading groups are formed during the fall of the first year. Reading groups [...] in general consist of 3-5 other students with whom you solve assignments, discuss the syllabus, exchange notes, etc. [...]." Further, it continues, "a good relationship within your group often has a very direct practical impact, for example, when forming reading groups." and "in addition, studying, and collaborating on home assignments and the like, can be more enjoyable and beneficial when the relationship is good".
Teaching were standardized across the different groups, and students faced the same exams and the same requirement.

The semester had 7 mandatory courses and each week had around 20 classroom hours. The number of weekly classroom hours declined an bit for the later semesters to around 15 hours. Students were expected to spend 40 hours a week in total on their studies.
The amount of in-class teaching was relative stable over the period we study. In 1986, it was a total of 470 teaching hours in the first academic year, and in 1995 is was 459 teaching hours. In the following year we don't know the exact number but the total study requirement did not change.
Elective courses were a new teaching style that was implemented in the beginning of the 80s and only few electives were offered. In 1986 elective courses were offered for 150 of the teaching hours (around 25%) in the final part of the studies; starting from the 4th semester.

[5] The first year was exclusively based on this classroom teaching, while some of the courses during the second and third year were combined more classes.

4.2.2 Data Sources and Sample Selection

For this study, we combine administrative data from Copenhagen Business School with Danish administrative register data from Statistics Denmark.

Our information about peer group composition stems from official records by the CBS administration for students who were enrolled between 1986 and 2006, and therefore our sample consists of 21 full cohorts of Business Economics students. The data provided by CBS contains information on the students' matriculation and exmatriculation dates, exmatriculation reasons, high school GPA, high school degree, citizenship, gender, age and most notably information on the initial peer group assignment by the CBS administration. Throughout our analysis, we keep all student observations regardless of their graduation status.

We augment the internal CBS data with Danish register data. In particular, we receive extensive information on the background and demographic information about Danish residents (as age, gender, marital status, place of birth, current place of residence, education) and most importantly annual labour market information on the universe of firms and workers in Denmark from 1980 to 2016. Additionally, we use the register data to get information on individuals outside the CBS sample to characterize jobs (workplaces, firms, occupations, industries) where students from our sample work after graduation.

The primary register-based data source for this study is the Danish matched employer-employee data (IDA). The data contains labour market outcomes (employment, occupation, industry, wages) and identifiers of firms and workplaces. To identify firms, we use the tax identity of the employer. Throughout the paper, we will use the terms "firm" and "employer" interchangeably. The definition of a workplace is not directly tied to an employer but corresponds to a physical location, namely the location where employees work (e.g. their office or a plant). All occupations are defined on the 4-digit level of the DISCO classification[6], and industries - on the 6-digit level of the DB classification. We use the annual cross-section of jobs from IDA for the entire period of study (1980-2016). For the years 2008-2016, we have access to the monthly employer-employee register (BFL), which we use to construct variables for labour market outcomes in this time-period (wages, hours and days of employment, industry and occupation codes). To study the educational trajectories of students in our sample, we use administrative data on individual education spells in Denmark (KOTRE).

Since this study focuses on studying students' labour market careers, certain sample restrictions apply. First, we do not observe student careers outside Denmark. For example, international students leaving Denmark after their studies or Danish students' careers abroad are not covered by our data. Second, as our study is focused on labour market networks, we do not consider observations outside of wage employment, which we define through the existence of an employment observation with non-zero wages and a non-missing employer identifier for a given year. As a consequence, we exclude

[6]When studying occupational similarities, we limit our sample to years with available occupational data (1994-2016) and consider only observations with non-imputed occupational codes.

observations of workers in both non-employment and self-employment. Below, we check if the sample restriction presents a threat to the empirical strategy.

4.2.3 Estimation Sample

We restrict our sample to the students' first spell in the Business Economics program. The sample includes 12716 students covering the cohorts from 1986-2006, and 360 randomized peer groups.[7] For more details see our companion paper Fischer et al., 2021.

For the main analysis, we construct a career panel that includes labour market information on the first 15 years after scheduled graduation (and hence 4-19 years after matriculation).

4.2.4 Summary Statistics

Table 1 shows summary statistics for the background characteristics of male and female students. Almost two-thirds of the students in our sample are male. On average, the students in our sample are older than 21 when they start their first spell at CBS. Few students (3.5%) are foreign citizens. High school GPA is measured in standard deviations from the distribution of a student's high school graduation cohort. On average, students have slightly higher grades than their high school cohort, and the female students in our sample have significantly higher high school grades than male students. An average student has one gap year after high school graduation and before the program start. Three-quarters of students start with having at least some work experience in Denmark. On average, they are from remarkably wealthy backgrounds: an average student's father is ranked in the 87th income percentile of the entire Danish income distribution. Strikingly, around 20% of fathers locate in the top 1% of the income distribution, while 20% of mothers are placed in the top 10%. For an average student, we observe around 35 peer group peers and 600 cohort peers.

Table 2 presents the ten most prominent occupations and industries for students in our sample. Students work in a wide array of white-collar occupations (finance, sales, administrative, managerial) and industries (finance, consultancy). Jobs in finance are dominating both in terms of industries and occupations.

How common is it for students from the same program/cohort/group[8] to share the same occupation, industry, employer or workplace? Table 3 presents how many students share a labour market "cell" with a group, cohort and program peer in a given year. Most importantly, it is not a rare event that workers in our sample share common career states with their peers. Non-surprisingly, the "broader" the labour market "cell", the more widespread is "working together". For example, 36% of group peers ever work in the same firm at a given point of time in their careers, while twice as many (73%)

[7]For some years of our sample period, CBS constructed specific classes for student with a background in marketing. They were not part of the randomization process and is excluded in our sample

[8]In this table cohort peers are also program peers and group peers are both - program and cohort peers.

work at least once in the same industry as their group peers. Secondly, as there are more cohort peers than group peers (groups are nested), it is much more common to be observed working with your cohort peer. 21% of students shared a workplace with their group peer, while almost 68% of students shared a workplace with a cohort peer.

Table 4 shows summary statistics for career outcomes in the career panel. Despite striking gender differences, both male and female students in our sample spend their careers at the top of the national income distribution. More than half of all observations are in the top 10% and a considerable share in top 1% of the income distribution in Denmark in a given year.[9] More than 10% of spells in the career panel correspond to a student working in a management position and 2.3% to spells in top management positions. Importantly, students tend to work in large firms.

[9]This observation is particularly remarkable given the fact that the sample includes early career observations and observations of students who dropped out of the program.

Table 1: Descriptive Statistics: Background Variables

	Male	Female	Total
Danish	0.966	0.962	0.965
	(0.181)	(0.191)	(0.185)
Age	21.47	21.40	21.45
	(2.291)	(2.394)	(2.327)
HS GPA	0.0931	0.223	0.138
	(0.792)	(0.718)	(0.770)
Gap years	0.988	0.958	0.978
	(1.346)	(1.354)	(1.349)
Work experience	0.746	0.745	0.746
	(0.435)	(0.436)	(0.435)
Father's education	13.65	13.58	13.62
	(2.779)	(2.806)	(2.788)
Mother's education	13.01	12.67	12.89
	(2.687)	(2.742)	(2.711)
Father's income rank	86.63	87.12	86.80
	(19.03)	(18.57)	(18.88)
Mother's income rank	70.90	70.45	70.75
	(22.52)	(22.52)	(22.52)
Father in top 1%	0.206	0.187	0.200
	(0.404)	(0.390)	(0.400)
Mother in top 10%	0.212	0.212	0.212
	(0.409)	(0.409)	(0.409)
Group size	36.11	35.35	35.85
	(5.793)	(5.989)	(5.872)
Cohort size	598.9	602.2	600.1
	(50.92)	(48.88)	(50.25)
Observations	8,218	4,300	12,518

Notes: This table shows mean values for individual background characteristics of students in the sample. Standard deviations are in parenthesis. Parental income is measured the year before student's matriculation.

Table 2: Descriptive Statistics: Top 10 Most Common Occupations & Industries

Occupation	Share	Industry	Share
Finance Professionals	14.12	Monetary intermediation	7.494
Financial and Mathematical Associate Professionals	6.937	Computer programming, consultancy and related	5.882
Legal and Related Associate Professionals	5.084	Accounting, bookkeeping, auditing, tax consultancy	4.962
Sales, Marketing and Public Relations Professionals	5.048	Public Administration, the economic and social policy	4.081
General Office Clerks	4.985	Wholesale of household goods	3.988
Sales and Purchasing Agents and Brokers	4.501	Advertising	3.397
Administration Professionals	4.414	Management consultancy activities	3.265
Software and Applications Developers and Analysts	3.710	Other financial service activities	2.498
Sales, Marketing and Development Managers	3.566	Secondary education	2.277
Business Services and Administration Managers	3.474	Pharmaceuticals	2.162
Total	55.84	Total	40.00

Notes: The table shows industry and occupation shares. Occupations are defined on the 3-digit level using the DISCO-08 classification. Industries are defined on the 3-digit level using the DB07 classification. Shares are calculated on all observations of careers in our sample for the years 2008-2016.

Table 3: Descriptive Statistics: Career Similarities

	Group Peers	Cohort Peers	Program Peers
Same workplace	0.210	0.677	0.911
	(0.0451)	(0.275)	(0.640)
Same firm	0.363	0.817	0.956
	(0.0849)	(0.398)	(0.733)
Same industry	0.731	0.985	0.999
	(0.279)	(0.810)	(0.986)
Same occupation	0.871	0.988	0.999
	(0.417)	(0.896)	(0.992)

Notes: The table reports share of students who have ever worked together with one of their group/cohort/program peers. Corresponding share of student-years in parenthesis. Industries are defined at the 6-digit level, occupations - at the 4-digit level.

Table 4: Descriptive Statistics: Career Panel

	Male	Female	Total
Income rank	85.81	83.20	84.87
	(17.69)	(16.41)	(17.28)
Top 10%	0.582	0.441	0.531
	(0.493)	(0.496)	(0.499)
Top 1%	0.102	0.0281	0.0751
	(0.302)	(0.165)	(0.263)
Log earnings	12.93	12.71	12.85
	(0.865)	(0.781)	(0.842)
Log daily wage	5.718	5.548	5.657
	(0.525)	(0.390)	(0.487)
Manager	0.127	0.0737	0.107
	(0.333)	(0.261)	(0.310)
Top manager	0.0305	0.0101	0.0229
	(0.172)	(0.1000)	(0.150)
Firm size (FTE)	377.4	444.5	401.4
	(667.7)	(908.4)	(763.4)
Observations	91,880	51,691	143,571

Notes: This table shows mean values for career outcomes in a panel of all available observations for employed students (from the first year after potential graduation). Standard deviations are in parenthesis.

4.3 Career Similarities & Networks

4.3.1 Identifying "Excess" Peer Similarities

To identify the effect of social interactions among university peers on their careers after graduation, we follow a dyadic approach and start with constructing all unique pairs of students (i, j) within each matriculation cohort $c(i, j)$.[10] In this study, we aim to identify the effect of interaction with group peers "in excess" of cohort peers. To illustrate the identification problem and the mechanics of dyadic regressions, let us consider the propensity of a pair of students (i, j) to work at the same job k at time t.[11] Observing a pair of students in the same job is not indicative of peer interactions *per se*. Students from the same cohort are likely to have similar abilities and career goals even before they

[10]We construct pairs as undirected dyads - (i, j) is equivalent to (j, i).
[11]Here, a "job" could be interpreted as an industry, occupation, firm or a workplace.

start their studies, which in consequence should lead to similar careers. Moreover, yearly variation in program contents might lead students from the same cohort to have more similar skill sets *ex post*. In the end, starting careers under similar (macro-) economic conditions could itself drive similarities in career paths between students. Therefore, the propensity of students to be observed at the same job could be expressed as

$$F_{ijkt} = \lambda_{c(i,j)kt} + \alpha_k P_{ij} + u_{ijkt},$$

where F_{ijkt} is the propensity of a pair (i, j) to work in the same job k at time t, $\lambda_{c(i,j)kt}$ reflects all factors that shape similarities in career choices of students from the same matriculation cohort $c(i, j)$ and P_{ij} measures the intensity of social interactions between two students.

Adding up the propensities for all (mutually exclusive) jobs $k \in K$:

$$F_{ijt} = \lambda_{c(i,j)t} + \alpha P_{ij} + u_{ijt}, \tag{4.1}$$

where α represents a theoretical parameter of interest. Estimation of Eq. 4.1 faces two major challenges. First, the intensity of social interactions P_{ij} is usually unobserved. Second, as students might choose to interact with someone who is *ex ante* more similar to themselves (a phenomenon which is often referred to as network homophily), it is likely that P_{ij} and u_{ijt} are correlated and the parameter of interest α would be unidentified. We use the peer group assignment as a proxy for the intensity in social interactions between students. To estimate the effect of peer interactions on career outcomes, we leverage both the fact that students are randomly allocated to their peer groups, and that the structure of the Business Economics program creates variation in time spent with group peers vs cohort peers.

Our approach of using peer group assignment to identify the effect of social interactions, therefore, relies on two assumptions. First, due to the random assignment, a given pair is equally likely to end up in the same group as any other pair of students within a given cohort. The random assignment solves the "selection problem" (Manski, 1993).[12] We provide evidence in favour of exogeneity of group assignment below. A second assumption is that - conditional on knowing the "true" intensity of social interactions - the peer group assignment is redundant for predicting F_{ijt}. In other words, we assume that the group assignment does not contain any useful information about the determinants of students' careers besides the compositions of their peer groups. The existence of group-level common shocks - factors that affect all group peers but are not the result of peer interactions *per se* - would violate this assumption. We assume that the uniformity of curriculum for a given course across the peer groups provides that the condition is likely to be satisfied. Besides group peers (the effect of which we are interested in), students in the same peer group are exposed to the same teacher assistants (TAs). Even though being assigned to the same peer group effectively means having the same TAs for mandatory courses, the supporting role of TAs and highly standardized

[12] Note that, since students were assigned randomly *conditionally* on gender, age and Danish citizenship, we always compare pairs of students conditioning on these covariates.

character of peer teaching process suggests that it is unlikely that they would have a strong and persistent effect on students' workplace choice. As we discuss below, we believe that the effect of TAs is unlikely to be consistent with the overall pattern of our findings. Previous research on the role of TAs suggest that they are unconnected with students' future academic outcomes (Feld et al., 2020). Moreover, we believe that using within cohort peer group randomization is an improvement over studies using between cohort comparisons. We expect that study conditions are much more similar for students from the same cohort than for students from two different cohorts. Provided that both assumptions are not violated, any "excess" similarities in careers between group peers over cohort peers can be attributed to the excess interaction within peer groups.

The intuition is formalized in the following simple empirical specification:

$$F_{ijt} = \lambda_{c(i,j)t} + \beta I_{ij} + \gamma X_{ij} + \epsilon_{ijt}, \tag{4.2}$$

where F_{ijt} is an indicator variable for an event of "working together" for a pair of students i and j in the year t; I_{ij} is an indicator of being assigned to the same peer group at the time of matriculation; $\lambda_{c(i,j)t}$ is a set of cohort/year/year-after-graduation fixed effects;[13] X_{ij} is a vector of dyadic covariates based on students' gender, age and status as a Danish citizen at the time of matriculation.[14] The causal coefficient of interest is β. In all dyadic regressions, we cluster on the level of randomization, the cohort. To address a (potential) threat to conducting inference due to a small number of clusters, we implement a wild cluster bootstrap.

Three important points need to be made concerning the interpretation of our parameter of interest, β. First, under the discussed assumptions, the proposed method allows us to identify not the total effect of group peers, but the effect "in excess" of the influence of cohort peers. P_{ij}, the actual level of interaction of a pair of students, should not be expected to be zero for cohort peers. Nevertheless, although cohort peers might be influential in shaping labour market outcomes as well, we assume that group peers interact more and are, on average, more influential. A positive and statistically significant estimate of β would provide evidence for more intensive interactions with peer group peers than with cohort peers. However, excluding the extreme case where students interact only within peer groups, this estimate is only a lower bound for the total effect of group peers. Second, since we do not observe the real network of social interactions between students, we effectively estimate an intention-to-treat effect. Not only is it realistic to assume that students interact between groups, but it is also likely to be true that not all students interact with the same intensity and quality with all of

[13] Hereafter, by the number of years after graduation, we mean the number of years after "scheduled" graduation - hence the number of years after matriculation plus the duration of the program (3 years). Therefore, the number of years after graduation becomes a predetermined, "mechanical" concept in our setting, which helps us to address the endogeneity of graduation timing.

[14] For each pair of students, a gender variable is defined specifying whether both students are male, female or differ in gender. Similarly, we define a variable specifying whether both students at the time of matriculation are Danish citizens, foreign citizens or mixed. Since age is a continuous variable, we follow a method described by Fafchamps and Gubert, 2007 for undirected dyads and calculate the absolute age difference within a student pair, the age sum and the squares of both terms.

their group peers. Moreover, some students drop-out and some students change groups. In none of our estimates do we condition on graduation or staying in the same group throughout the studies. Therefore, β measures the effect of being initially assigned to the same peer group. Last, since both, F_{ijt} and I_{ij}, are indicator variables, β is a percentage point difference between frequencies. Given that the magnitude of these differences is hardly intuitive in the dyadic setting, we also calculate the effect in percents relative to a baseline measure of similarity. For example, if F_{ijt} equals to 1 for a pair of students (i, j) having the same occupation at year t, then a pair of students from the same peer group is β percentage points more likely to be observed with the same occupation than a pair from the same cohort but different groups. To get a better understanding of the magnitude, we divide β by the baseline frequency for students from the same cohort but different groups.

4.3.2 Evaluation of the Empirical Strategy

The conditional random assignment of students to peer group peers is crucial for the identification of the influence of social interactions on career dynamics. In the absence of random assignment, if students can choose their peers, a selection problem arises (Manski, 1993). In this case, observed peer similarities in labour market outcomes are driven by initial similarities between students that are unobservable to the econometrician. As our identification strategy crucially depends on the conditional random allocation of students to peer groups, we demonstrate the balancedness of our sample by showing that group peers are not initially more similar than cohort peers.

We implement a version of the balancing test based on Eq. 4.2, where future career states are replaced with a set of predetermined variables:

$$F_{ij} = \lambda_{c(i,j)} + \beta I_{ij} + \gamma X_{ij} + \epsilon_{ij}, \tag{4.3}$$

where I_{ij} is an indicator of being assigned to the same peer group at the time of matriculation and $\lambda_{c(i,j)}$ are matriculation cohort fixed effects. F_{ij}, in this case, might be both an indicator variable reflecting that students belong to the same category or absolute difference between values of some predetermined variables for a pair of students (i, j).

For the balancing test, we use 12 characteristics measured at the time of matriculation - educational background (high school GPA, high school track, non-high school degree finished), municipality of residence, family background (mother's and father's number of years of education, mother's and father's disposable income rank) and previous employment histories (previous workplaces, firms, occupations and industries). As it is shown in Table 5, none of the variables appear to be unbalanced on the 5% level. Concerning the municipality of residence before matriculation, test indicates marginally significant on the 10% level sorting, which could be due to the number of tests performed. Therefore, we conclude that the sample is balanced on predetermined similarities between group and cohort peers.

Table 5: Dyadic Balancing Test

	GPA	HS Track	Non-HS Degree	Municipality
Same group	0.000529	0.0159	-0.00169	0.110*
	(0.766)	(0.905)	(0.317)	(0.0900)
Observations	3,308,941	3,166,697	3,749,507	3,749,507
	Mother's Education	Father's Education	Mother's Income	Father's Income
Same group	0.00382	-0.00266	-0.0329	-0.0335
	(0.422)	(0.615)	(0.444)	(0.550)
Observations	3,075,643	2,795,315	3,318,792	3,088,727
	Workplace	Firm	Industry	Occupation
Same group	0.0178	0.0236	0.00770	0.0391
	(0.122)	(0.426)	(0.869)	(0.165)
Observations	3,749,507	3,749,507	3,749,507	3,749,507

Notes: The table reports the balancing test as specified in Eq. 4.3 for the following variables measuring predetermined similarities between students: the difference in high school GPA, the same high school track, both students with non-high school degree, the same municipality of residence, the difference in years of education of mothers, the difference in years of education of fathers, the difference in disposable income ranks of mothers, the difference in disposable income ranks of fathers, worked at the same workplace, firm, industry and occupation five years before matriculation. *** p<0.01, ** p<0.05, * p<0.1

4.3.3 Results

4.3.3.1 Main Results

We start with estimating the "excess" similarities formalized in Eq. 4.2. Table 6 explores the "excess" similarities between group peers in their choice of industry, occupation, firm and workplace. For all four outcome variables, we find that pairs of group peers are more likely to "intersect" in their career paths than pairs of cohort peers.

Since, as it was discussed above, point estimates measured in percentage points do not help to gain an intuition for both magnitude and relative importance of the effects, we also report the effect in percents relative to a baseline, which we define as the propensity of cohort peers to be observed working together at the same "cell". The observed similarity of point estimates for all outcomes hides large differences in effects measured in comparison to the baseline. A pair of students is around 4% more likely to work in the same industry and the same occupation if they were initially assigned to the same peer group. However, the effect is much larger when less aggregated labour market "cells" are considered. For a pair of students, being allocated to the same peer group leads to a 23% higher probability to work at the same firm and an increase by

Table 6: Career Similarities: Baseline Regressions

	Same Industry	Same Occupation	Same Firm	Same Workplace
Same group	0.0858***	0.136***	0.0877***	0.0779***
	(0.00660)	(0.000800)	(0)	(0)
Effect(in %)	3.999	3.967	23.09	40.16
Baseline	2.146	3.430	0.380	0.194
R-squared	0.00188	0.00110	0.000356	0.000446
Observations	20,707,850	12,201,751	25,073,724	20,419,748

Notes: This table shows estimates of the linear probability model as specified in Eq. 4.2. Coefficients and baselines are multiplied by 100 to reflect percentage points. P-values calculated using wild cluster bootstrap (5000 replications) on matriculation cohort level are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

40% in the probability to work at the same workplace after the graduation.[15]

So, we see that the effect of peers is concentrated at the most disaggregated level - the workplace. One might find this to be at odds with both, the idea that peer interactions are restricted to the period before graduation only and common academic environment as a major factor behind the excess similarities. Human capital spillovers at the classroom level, co-formation of career preferences and influences of TAs are naturally expected to affect general career trajectories which would be proxied by industry and occupation choices.[16] On the other hand, active alumni networks might act as a source of job-related information, from which, for example, graduates gain knowledge about job-openings or refer each other to employers. This type of interaction is expected to go in line with the strong firm and workplace similarities.

To further corroborate this reasoning, we ask whether the workplace effect drives the effects on the occupation-, industry- and even firm-level. For instance, group peers may be more likely to choose similar industries just because they are, in the first place, more likely to work at the same workplaces. To answer this question, we redefine ”working together” as an event of ”working together (at the same firm, or in the same occupation or industry), but not at the same workplace”. As it is evident from Table 7, the effects of working in the same industry and firm lose their statistical significance after the exclusion of workplace similarities. Only for occupational choices, the effect stays significant on 10% level. We conclude that peer similarities in career outcomes are driven by interactions on the most desegregated level - the workplace. Furthermore, we interpret this result as suggestive evidence that students form post-graduation networks, and that post-graduation interaction in the labour market is a key driver of career similarities of former academic peers.

[15]The contrast between the point estimates and the recalculated relative effects is caused by the fact that even though the treatment leads to almost the same increase in frequencies of ”working together” events in percentage points, the baseline probability of being observed at the same workplace is much lower than the baseline probability of being observed at the same industry. The latter pattern is also reflected in Table 3.

[16]Otherwise, classroom interaction between students would need to affect human capital along firm-specific/workplace-specific lines and/or influence preferences towards specific firms/workplaces.

Table 7: Career Similarities: Net of Workplace Effects

	Same Industry	Same Occupation	Same Firm
Same group	0.00936	0.0791*	0.00976
	(0.722)	(0.0504)	(0.249)
Effect(in %)	0.474	2.345	4.598
Baseline	1.974	3.373	0.212
R-squared	0.00161	0.000967	0.000163
Observations	20,417,993	10,278,441	20,419,748

Notes: This table shows estimates of the linear probability model as specified in Eq. 4.2. Coefficients and baselines are multiplied by 100 to reflect percentage points. P-values calculated using wild cluster bootstrap (5000 replications) on matriculation cohort level are in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

4.3.3.2 Timing and Heterogeneity

Does the intensity of post-graduation interaction of former academic peers decrease over time? Figure 1 illustrates the timing of ”excess” firm and workplace similarities. In both cases, the effect is strongest a few years after scheduled graduation and diminishes over time.[17] For working for the same firm, the magnitude of the effect decreases more than twofold over time: from above 40% to less than 20%. At the start of their careers, a randomly chosen pair of group peers is almost twice as likely to share a workplace as a randomly chosen pair of cohort peers. Ten years after scheduled graduation, this effect decreases by four times but is still economically significant - 22%.

Social connections tend to form more intensely between individuals that are more similar (e.g., Eliason et al., 2019). This phenomenon is usually referred to as network homophily. In the context of social networks among potential top-earners, an increased propensity to interact with similar students might potentially exacerbate inequality and hamper social mobility for less advantaged groups. Therefore, it is of interest to investigate whether the strong peer effects in workplace choices are more pronounced for more similar students.

We observe the most striking evidence of homophily among gender lines (Figure 2). We apply the same empirical approach as before, but explore the timing of the effect separately for pairs of students of the same gender and pairs of students of a different gender. The observed pattern of peer ”excess” similarity is mostly driven by pairs of students of the same gender. For different gender pairs, the effect is significant only for the first two years after graduation and not significantly different from zero afterwards. On the other hand, for same-gender pairs, the effect is as high as 134% the year after graduation and still 45% ten years after graduation.

Figure 3 further investigates heterogeneous effects by gender, country of origin, age and high school GPA. First, we divide same-gender pairs of students into pairs of male students and pairs of female students. The effect for both types of same-gender dyads is significantly higher than for mixed dyads (for which the effect is not significantly

[17]This pattern is broadly consistent with findings in other social contexts (Eliason et al., 2019).

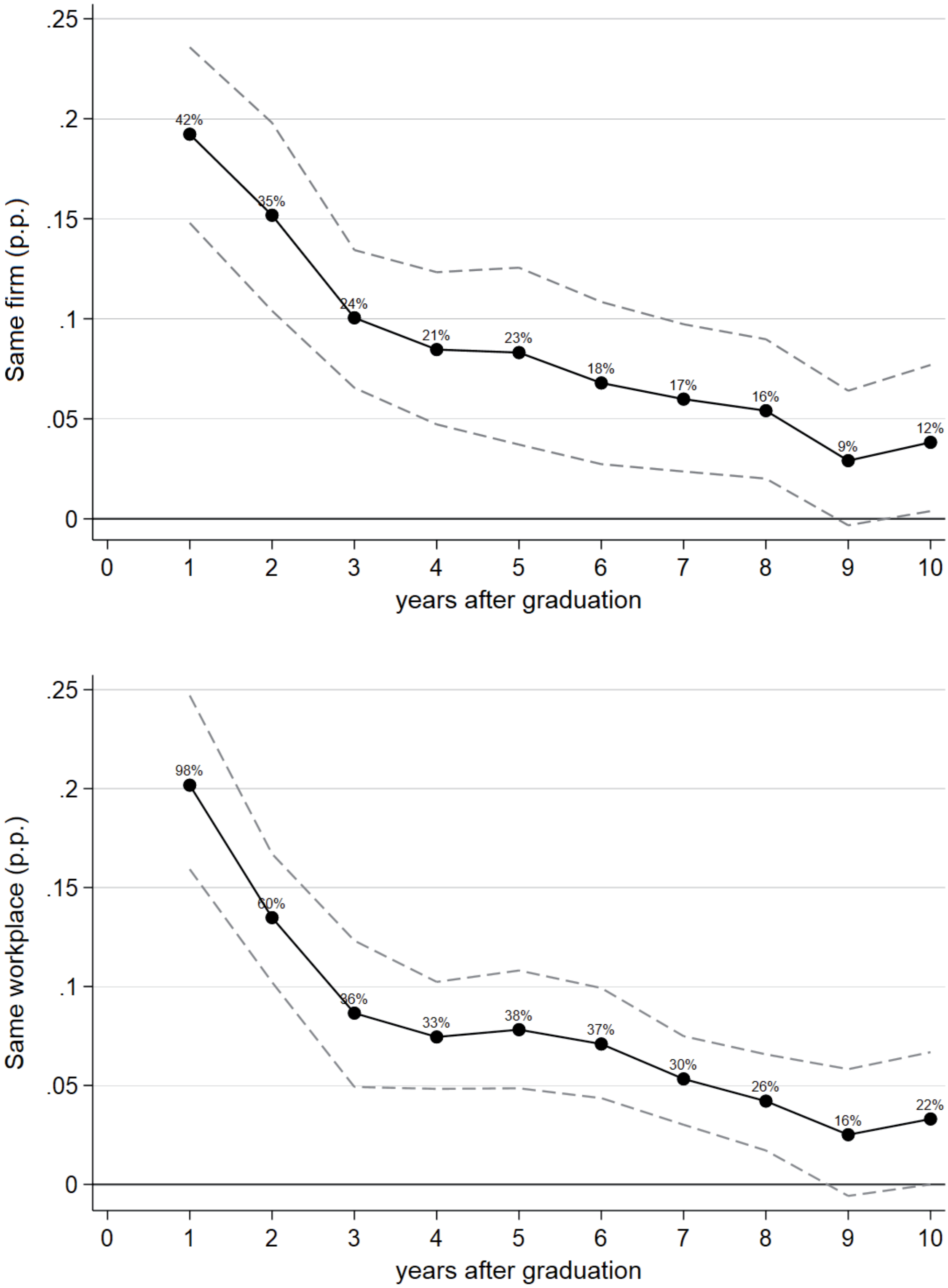


Figure 1: Same Firm and Same Workplace: Timing of the Effect

Notes: The y-axis represents the effect of being assigned to the same group in percentage points. The relative effects (in %) are represented on a graph. 5% confidence intervals refer to point estimates. Standard errors are calculated by wild cluster bootstrap at the matriculation cohort level (5000 replications)

.

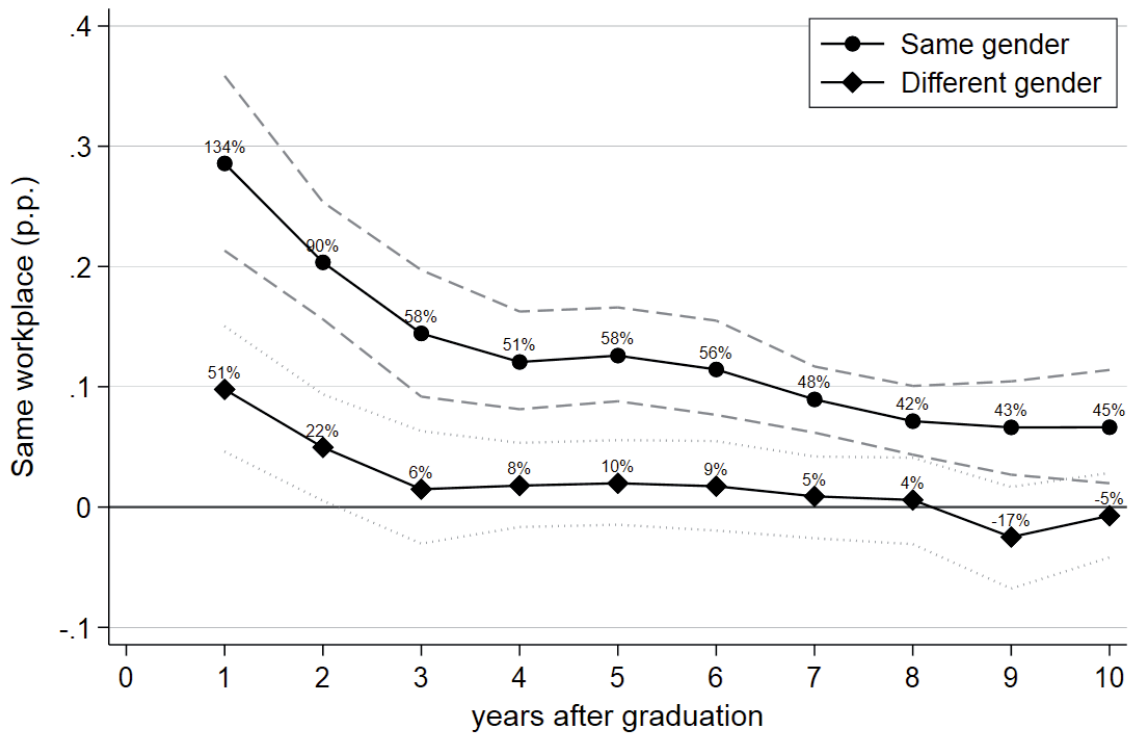


Figure 2: Same Workplace: Timing of the Effect, by Gender

Notes: The y-axis represents the effect of being assigned to the same group in percentage points. The relative effects (in %) are represented on a graph. 5% confidence intervals refer to point estimates. Standard errors are calculated by wild cluster bootstrap at the matriculation cohort level (5000 replications)

.

different from zero). Although the point estimate for male pairs is slightly higher, we cannot statistically distinguish the effect for female dyads from the effect on male dyads. Another dimension of (potentially) important heterogeneity is country of origin. We investigate if the effect for the student pairs of the same origin is different from the pairs of different origin. Only a few students in our sample are not Danish citizens and (by design) the dyad data construction drives the relative share of non-Danish dyads down. Hence, we cannot investigate Danish dyads separately from dyads of immigrants from the same source country.[18] Consequently, we compare the magnitude of the effect for pairs of students with the same country of origin (including Danes and non-Danes) to pairs of students with a different origin. As we see, the effect for same-origin pairs is significantly higher (on the 5%-level) than for different origin pairs, while the latter is

[18]The share of dyadic observations constructed from a given group of students is much smaller than the share of these students in the population. If there are n students of a given type in a cohort of size N, the share of this type in a cohort is $\frac{n}{N}$, but the share of dyads constructed from students of these type will be $\frac{n(n-1)}{N(N-1)}$.

not significantly different from zero. This finding suggests that the social interactions in our sample are characterized not only by gender homophily but also by country-of-origin homophily.

Furthermore, we look at heterogeneity by age at the time of matriculation and high school GPA. Since both variables are continuous, we implement a different approach by interacting the treatment with the absolute difference of matriculation age (GPA) and the sum of matriculation age (GPA) for a student pair.[19] These two types of interactions are meant to reflect the different characteristics of the effect. The difference-interaction-term is aimed to capture the homophily of the effect - namely if students of a similar age and with similar grades tend to interact more. On the other hand, the sum-interaction-term captures if the effect is different students for older students and students with higher GPA. As we see, neither the age difference nor the age sum interaction appears to be statistically significant. The only significant interaction for these variables is GPA-sum. For a pair of students with one standard deviation higher sum of high school GPA, the effect of being assigned to one peer group is almost twice as high. This finding suggests that more able students are more likely to make use of alumni networks.

As we have already highlighted, not only does the typical student from our sample end up having a career at the top segment of the Danish labour market but also most of the students stem from affluent families. In a related study, S. D. Zimmerman, 2019 suggests that a higher tendency of students from a wealthy background to form active labour market networks after graduation might explain why the returns to "elite" education programs are unequally distributed between students with different social backgrounds. In short, educational returns might be higher for students from a wealthy background because they accumulate more social capital.

To address this question, we contrast the effects of being assigned to the same peer group on the probability of working at the same workplace for specific dyads which we classify by the fathers' disposable income rank in the year before matriculation (Figure 4). We perform three comparisons, where we are defining "rich" dyads as pairs of students where both have fathers in the top 1%, top 10% and top 25% of the disposable income distribution in Denmark and contrast these effect to the remaining pairs of students (hence including the pairs where one of the students has a father in top income group). Even though the classifications based on fathers in the top 10% and top 25% lead to higher point estimates for the "rich" dyads, the difference with other dyads is not statistically significant. However, looking at students with fathers in top 1% shows that effects for this group of students is more than twice as large (as compared to all other students) and the difference is statistically significant at the 5% level.[20] Therefore, we conclude that, indeed, the strongest networks tend to emerge among the students coming from the richest families.

[19]For the main effects to be comparable to the baseline estimate in Table 6 we standardize the absolute differences and sums within our sample.

[20]Looking at pairs of students with fathers at the very top of disposable income distribution leads to a problem similar that we face with non-Danish students. Even though one-fifth of all students have fathers in that group, a much lower fraction of resulting dyads has both students in that group (around 4%). Consequently, at the very top, our estimates are noisier.

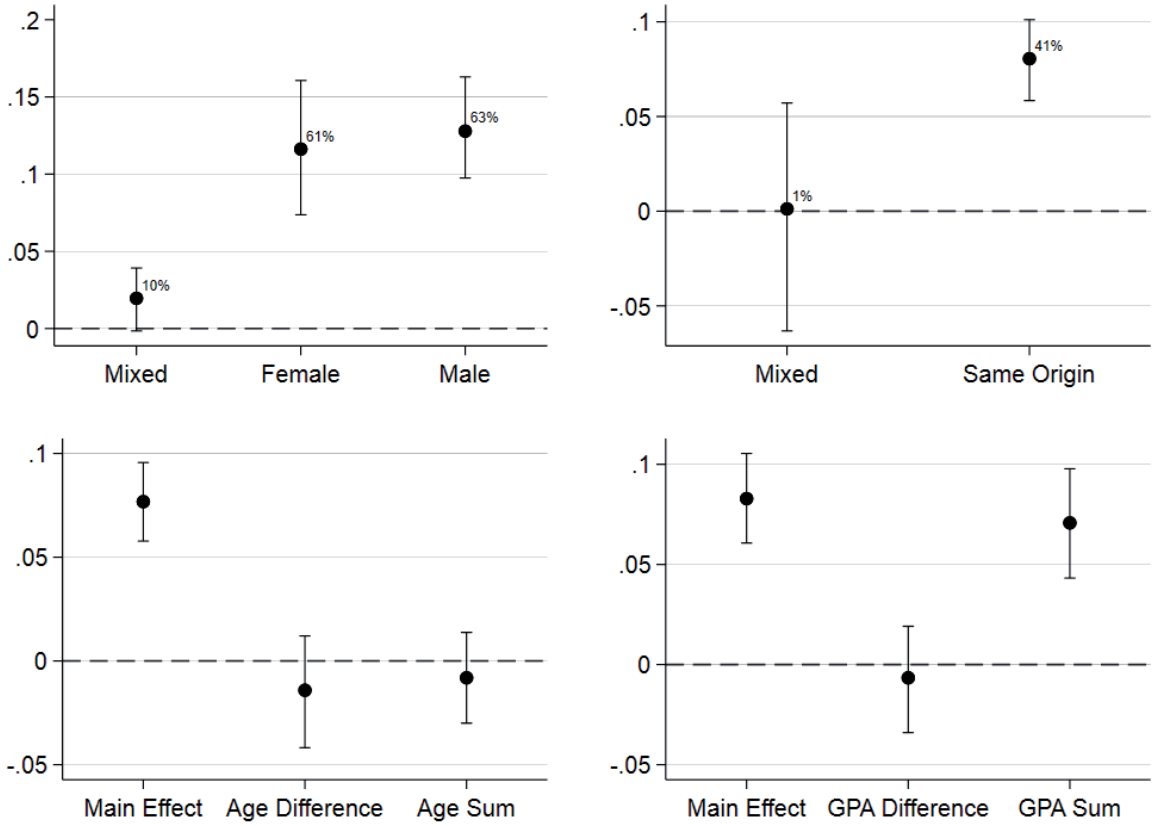


Figure 3: Same Workplace: by Gender, Country of Origin, Age and High School GPA

Notes: The y-axis represents the effect of being assigned to the same group in percentage points. The relative effects (in %) are represented on a graph. 5% confidence intervals refer to point estimates. Standard errors are calculated by wild cluster bootstrap at the matriculation cohort level (5000 replications)

.

4.3.3.3 Robustness Checks

In this section, we address several threats to our interpretation of the results. If peer interaction in class leads students to either work abroad or to be self-employed, our estimates might be biased. We, therefore, check if selection out of our sample might invalidate our results. Next, we ask if subsequent education choices mediate the observed career similarities. The observed similarities of peers in careers would then not be the result of labour market interaction, but at least partially a direct result of similar education choices. Last, we investigate if the linearity imposed by our baseline specification - a linear probability model - leads to the quantitatively and qualitatively deceptive results.

Even though the evidence presented in Table 5 supports the assumption of (conditionally) random assignment of students to peer groups in our sample, not all of the students are observed at each year in our career sample. Therefore, for some student pairs, some dyad-years are missing. There are different reasons why a student is not observed in

Table 8: Career Similarities: Sample Selection

	Sample (All)	Sample (5 years)	Population Sample (All)	Population Sample (5 years)	Employment Sample (All)	Employment Sample (5 years)
Same group	0.0447	0.0293	0.0330	0.00759	0.0215	0.0260
	(0.268)	(0.497)	(0.322)	(0.689)	(0.542)	(0.494)
Baseline	67.22	68.55	89.19	95.74	75.35	71.58
R-squared	0.0566	0.0565	0.106	0.0978	0.0250	0.0371
Observations	64,845,094	18,747,535	64,845,094	18,747,535	57,850,264	17,948,399

Notes: The table reports results of the sample selection tests (baseline specification of Eq. 4.2 is used). Outcome variables are indicator variables equal to one if both students are observed in our career sample in a given year (”Sample”), in population of Danish residents in a given year (”Population Sample”) or in the career sample conditional on both being observed as Danish residents in a given year (”Employment Sample”). For each outcome variable we use both all available observations and only first 5 years after (scheduled) graduation from the program. Coefficients and baselines are multiplied by 100 to represent percentage points. P-values in parenthesis are calculated by wild cluster bootstrap on matriculation cohort level (5000 replications). *** p<0.01, ** p<0.05, * p<0.1

Table 9: Career Similarities: Educational Choices

	Graduate from CBS HA	Graduate from any HA	Graduate from any Bachelor's	Switch to other program
Same group	0.0187	0.0185	0.0755	0.0170
	(0.831)	(0.844)	(0.383)	(0.192)
Effect(in %)	0.0382	0.0375	0.137	2.411
Baseline	49.01	49.33	55.08	0.703
R-squared	0.0481	0.0488	0.0582	0.00272
Observations	3,428,983	3,428,983	3,428,983	3,428,983

	Master's start	Master's graduate	Master's program	Master's institution
Same group	0.0418	0.00417	0.238	0.0484
	(0.688)	(0.952)	(0.193)	(0.494)
Effect(in %)	0.0783	0.0133	0.409	0.0602
Baseline	53.36	31.39	58.24	80.26
R-squared	0.0615	0.0629	0.00908	0.0217
Observations	3,428,983	3,428,983	1,830,807	1,830,807

Notes: This table shows estimates from the linear probability model as specified in Eq. 4.3 for indicator variables equal to 1 if both students graduate from the program, graduate from any Business Economics program, graduate from any Bachelor's program, switch to another program, start any Master's program, graduate from any Master's program, start the same Master's program, start any Master's program at the same institution. Coefficients and baselines are multiplied by 100 to reflect percentage points. P-values calculated using wild cluster bootstrap (5000 replications) on matriculation cohort level are in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

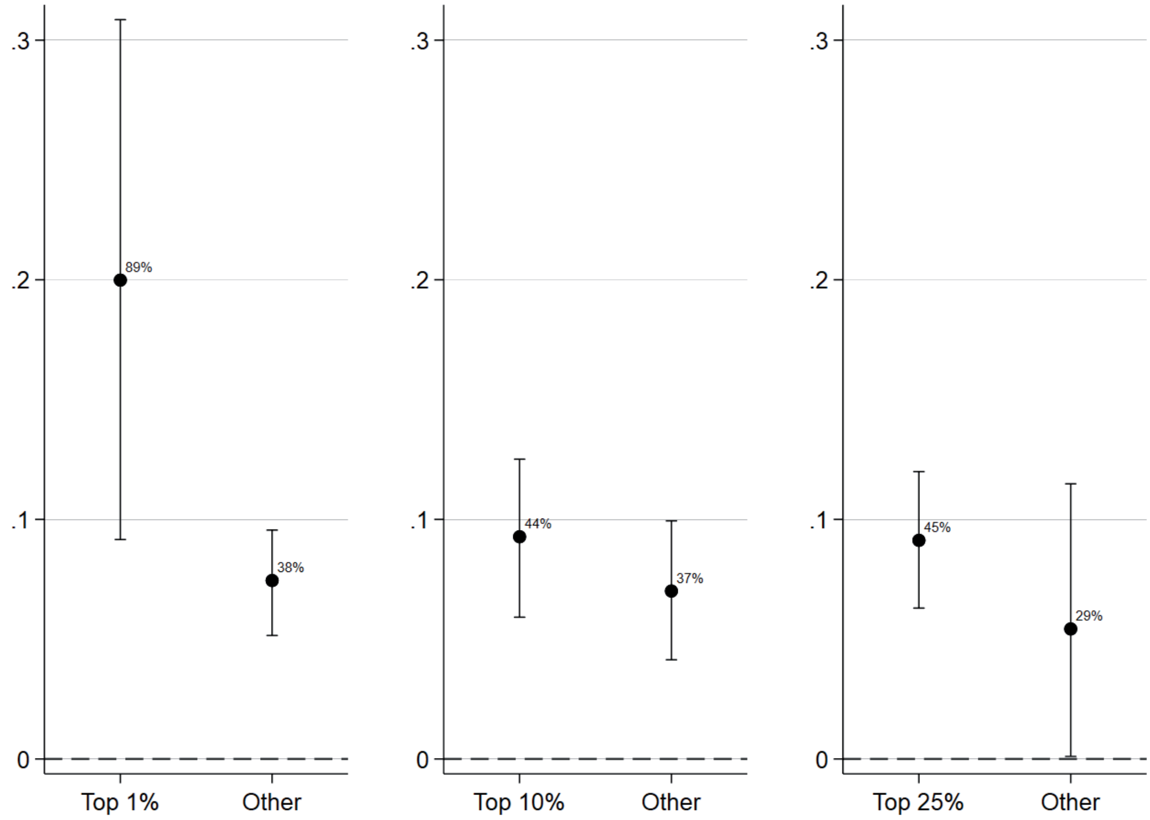


Figure 4: Same Workplace: by Father's Income Rank

Notes: The y-axis represents the effect of being assigned to the same group in percentage points. The relative effects (in %) are represented on a graph. 5% confidence intervals refer to point estimates. Standard errors are calculated by wild cluster bootstrap at the matriculation cohort level (5000 replications)

.

our sample in a given year. First, some students might leave Denmark temporarily or permanently (this is particularly relevant for international students). Second, even students who stay in Denmark might be out of wage employment (e.g., non-employed or self-employed). If the peer group assignment process affects students' decision to leave the sample, for example, through peer effects in migration, and that decision is correlated with the initial propensity of students to make similar career choices, then missing dyads might cause our estimates to be biased.

To check whether sample selection could potentially cause bias in our estimates, we estimate the effect of being assigned to the same peer group on being both observed in the wage employment sample in a given year (we use the same specification as for the baseline Eq. 4.2). The first two columns of Table 8 present results of the test using, first, all available observations in the data and, second, only the first five years after planned graduation from the program (which is where we observe the most substantial effects for our main results). As we see, group peers are not significantly more likely to be both missing from the sample in comparison to cohort peers (neither overall nor in the first five years). We further investigate if group assignment affects any of the stages of the selection process: both being observed as Danish residents (”Population Sample”) and both being observed in our career sample *conditional* on being both observed among Danish residents (”Employment Sample”). Neither of the tests identifies significant correlations with the group assignment. We, therefore, conclude that sample selection is unlikely to cause bias to our estimates.

Although in this study, we investigate the effects of peer interactions on careers, similarities in career choices may arise because peers affect each other's educational choices that precede their career start. For example, if peers tend to have similar drop-out behaviour, this might result in similar jobs even without further post-graduation interaction in the labour market.[21] Table 9 shows regression results for the specification of Eq. 4.3 to similarities in educational choices on a Bachelor's level (graduation from the CBS Business Economics program, graduation from any Business Economics program, graduation from any Bachelor's program, switch to a different program) and a Master's level (start of any Master's program, graduate from any Master's program, start the same Master's program, start any Master's program at the same institution). We do not observe any significant effects of peer group assignment on educational choices. Hence, this channel is unlikely to explain our main findings.

Last, we check if using the linear probability model in our baseline specification (Table 6) provides misleading results. Given that the outcome variable is distributed highly unevenly, the linear model might provide a bad approximation. We repeat our baseline analysis using a logit specification. As we can see, the coefficients in Table 10 have the same relative order and the same order of magnitude as the coefficients in our baseline specification - the strongest effects are again observed at the most disaggregated level. According to this specification, being assigned to the same group increases the probability for a pair to work together at the same workplace by 31.6%. The average marginal

[21] Note that, even in this case, it would be hard to explain how similarities in educational choices are driving career similarities at the workplace level. Similar educational choices would likely cause similarities at the level of occupation and/or industry choice.

Table 10: Career Similarities: Logit Specification

	Same Industry	Same Occupation	Same Firm	Same Workplace
Same group	3.549**	3.889***	20.25***	31.55***
	(0.0212)	(0.000800)	(0)	(0)
AME	0.0747	0.129	0.0777	0.0627
Baseline	2.146	3.430	0.380	0.194
Pseudo R-sq	0.00945	0.00377	0.00678	0.0163
Observations	20,707,850	12,201,751	25,073,724	20,419,748

Notes: This table shows estimates from the logit specification of a baseline regression (Table 6). Average marginal effects and baselines are multiplied by 100 to reflect percentage points. P-values calculated using score cluster bootstrap (5000 replications) on matriculation cohort level are in parenthesis (Kline & Santos, 2012). *** p<0.01, ** p<0.05, * p<0.1

effects from the logit specification mirror the coefficients in the linear probability model.

4.4 The Effect of Working Together

4.4.1 Empirical Strategy

In the previous section, we have shown that peer interaction leads to a tendency to share the same workplace with group peers. However, does joining a group peer also improve labour market outcomes? Does post-graduation interaction with peers benefit workers' careers?

A key challenge for identifying if joining the firm of an incumbent group peer is beneficial is the choice of a proper comparison group. For many reasons, comparing outcomes of "joiners" to outcomes of "non-joiners" does not help to answer the question at hand. First of all, workers joining their peers might ex-ante be different from workers not joining.[22] Second, even without social interactions, (voluntary) job-to-job transitions tend to be associated with wage increases. Moreover, employers in our sample are far from being average. Therefore, even by comparing the effect of joining a group peer's workplace to job-to-job transitions events does not disentangle the effect of "working together" from the effect of working at a specific type of workplace. Finally, workers who decide to join their peers may be subject to unobserved shocks, which would affect their careers even in the counterfactual scenario of not joining a group peer's workplace.

To address these concerns, we employ a Difference-in-Difference type strategy[23] and compare career trajectories of workers who join a workplace with one or more incumbent group peers with workers who join a workplace with one or more incumbent cohort peers. Using the event of joining cohort peers as a counterfactual resembles the "excess" similarity strategy discussed above. Since real social interactions are unobserved and may be active across cohort peers as well, the interpretation of the difference between joining a group peer and cohort peer depends on the assumption that the former is more likely to be the result of "networking" among former peers than the latter.

We use within-worker variation (individual fixed effects) to account for systematic differences between workers hired through social connections and "formal" channels. The comparison of group peer "joiners" to cohort peer "joiners" addresses the identification threat from the firm composition. As group peers for one student are cohort peers for another, there should be no ex-ante systematic difference in the general characteristics and "quality" of firms between establishments where group peers or cohort peers work. Although, this does not mean that it is necessarily the case *ex-post*. Firms, where

[22] Relying on previous research (e.g., Kramarz and Skans, 2014), one might expect workers who find jobs through their social connections to be negatively selected.

[23] In particular, the identifying variation that we use stems from a Difference-in-Difference type comparison. To accommodate statistical power concerns, resulting from the necessity to identify a set of fixed effects and controls, we use all information about student careers in our sample, not only on the events of joining a cohort or a group peer.

workers join their former group peers, might on average be better or worse than firms at which they join their cohort peers. Such an ex-post difference would point towards the fact that alumni networks are used to get access to specific jobs. Even though the existence of unobserved shocks (correlated with the decision to join a peer's workplace) is inherently untestable in our setting, we provide evidence in favour of common trends between students who join an incumbent peer-group member and students who join an incumbent cohort member below.

Our empirical strategy differs from the approaches that are commonly employed in the literature, which studies the effects of referrals on labour market outcomes. Some studies rely on detailed personnel records and compare workers hired through referrals to non-referred workers (for example, Brown et al., 2016 and Burks et al., 2015). Another branch of research deals with the selection problem of workers who are hired through referrals by using linked employer-employee data and employing both - worker and firm fixed effects (e.g., Dustmann et al., 2016 and Zhu, 2018). From the point of an employer, a referral wage premium (or penalty) represents a wage differential between otherwise similar workers that arises solely from the hiring channel of the worker. However, from the worker's perspective, this differential is not the only source of benefits from job search networks (referrals or information sharing about job openings). Workers could benefit from getting access to jobs in higher-paying firms (or occupations). Controlling for firm-fixed effects suppresses these additional benefits of labour market networks for the worker. As we aim to capture the full picture of potential benefits (or costs) of labour market networks, we do not control for firm-fixed effects.

Our approach is formalized in the following regression framework:

$$y_{ijt} = \lambda_t + \mu_i + \alpha * CohortInc_{ijt} + \beta * GroupInc_{ijt} + \gamma X_{it} + \epsilon_{ict}, \tag{4.4}$$

where y_{ict} is the outcome variable of interest for individual i at firm j in year t; λ_t is a year fixed effect; μ_i is an individual fixed effect; $CohortInc_{ijt}$ is an indicator which equals to 1 if at the time where worker i joins firm j there was at least one cohort peer (from the same or different group) working at firm j; $GroupInc_{ijt}$ - is an indicator which equals to 1 if at the time when worker i joins firm j there was at least one group peer working at firm j; X_{it} are age and experience third degree polynomials.

Note that β is the parameter of interest, which is identified through a comparison between events of joining a group peer versus a cohort peer, and captures the benefits (or costs) of finding a job through the alumni network. On the other hand, α does not have a similar interpretation, as it compares transitions to cohort peers to all other observations (including individuals not changing a firm).

4.4.2 Evaluation of the Empirical Strategy

First of all, the validity of the proposed empirical strategy depends on a parallel trend assumption. The potential outcomes of students who join a group peer should be similar to those who join a cohort peer. To provide suggestive evidence in favour of this assumption, we look for differences in the pre-trends between the treatment and control groups.

If students who join their former cohort peers as a comparison group for students who join their former group peers show similar career dynamics before the event of the job change, we interpret this as suggestive evidence that the assumption of parallel trends in potential outcomes holds.[24]

Table 11: Group Peers vs Cohort Peers

	Joiner	Non-Joiner
Cohort Wage	6.886	7.004
	(0.493)	(0.449)
Group Wage	6.885	7.004
	(0.513)	(0.466)
Cohort Occ. Pay	62.19	63.86
	(8.318)	(7.724)
Group Occ. Pay	62.19	63.86
	(9.632)	(9.144)
Cohort Industry Pay	58.01	58.62
	(4.406)	(3.884)
Group Industry Pay	57.97	58.61
	(7.681)	(7.169)
Cohort Firm Pay	64.19	64.55
	(4.666)	(4.095)
Group Firm Pay	64.14	64.55
	(7.908)	(7.344)
Cohort Firm Size	384.3	395.5
	(59.63)	(58.23)
Group Firm Size	382.6	395.9
	(158.2)	(161.4)
Observations	3,348	147,049

Notes: This table shows mean values and standard deviations for labour market outcomes of students' group and cohort peers - for students who join their peer's firm and all students in the sample. Occupation, industry and firm pay ranks are constructed from the occupation, industry and firm fixed effects from the multiway fixed daily wage regressions with individual fixed effects, year fixed effects and age polynomials on the whole population of Danish workers (1985-2016).

[24] Note that the same students could be potentially in treatment and control groups in their careers.

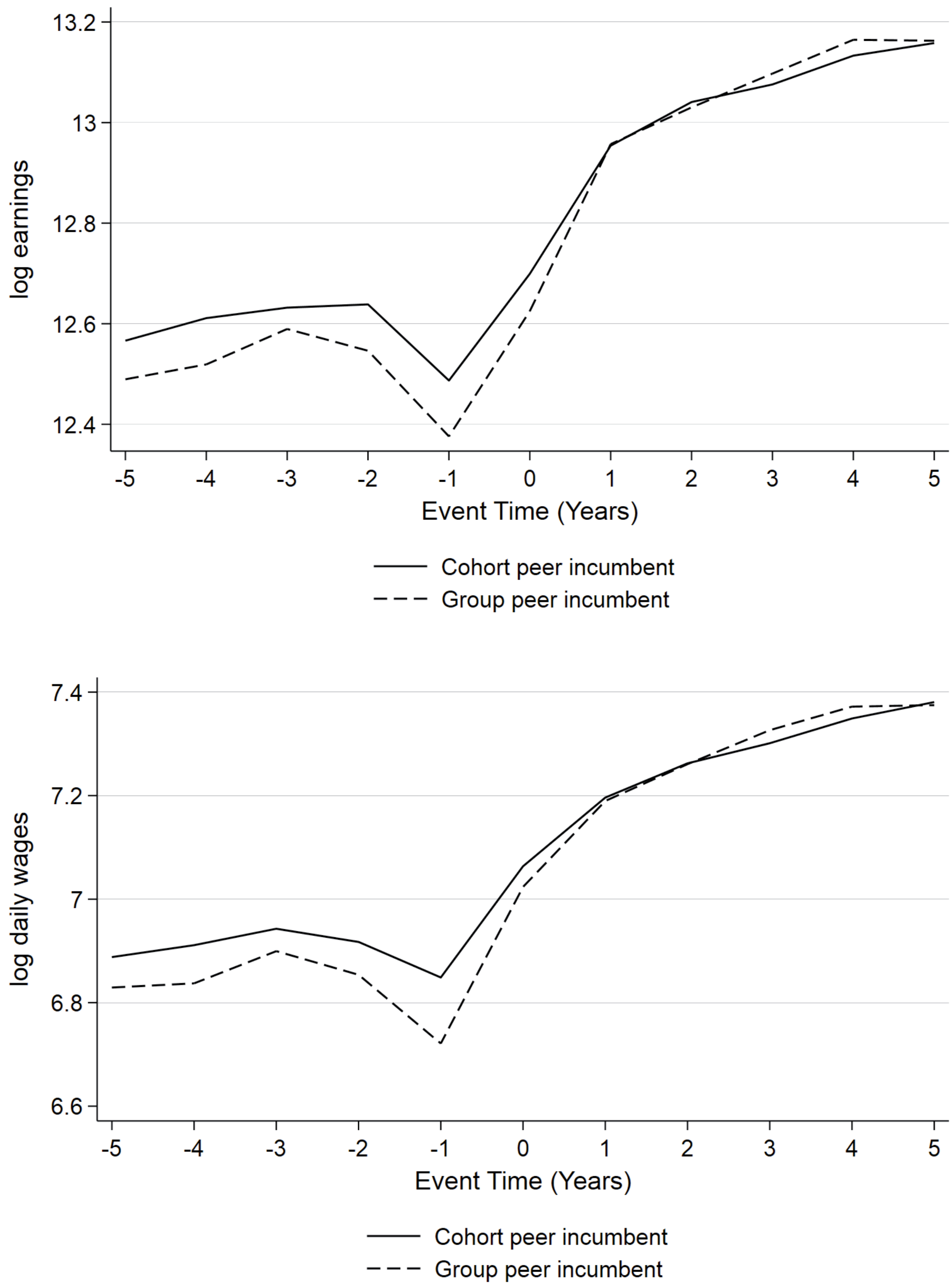


Figure 5: Wage Dynamics: Group Peer Joiner vs Cohort Peer Joiner

Notes: The y-axis represents real log annual/daily wages, the x-axis - number of years relative to the job-change event. ”Cohort peer incumbent” line depicts dynamics for students joining a firm where someone from their cohort works, ”Group peer incumbent” line - dynamics for students joining a firm where someone from their peer group works. Event time 0 corresponds to a year when a student joins a firm.

Figure 5 shows differences in dynamics between students who join a cohort peer and those who join a group peer for both earnings and daily wages. Most importantly, both groups show very similar wage dynamics before the job changing event.[25] However, after the event of a job change, workers joining their former group peers experience faster wage growth. Still, this gap is almost closed the year after the job change. Hence, we conclude that the treatment group is not selected on the pre-trend (e.g., initially faster-growing wages).

Another crucial component of our empirical approach is an assumption about the absence of *ex-ante* differences between group peers and cohort peers. For example, if there is no systematic difference between jobs where group peers and cohort peers work, but joining a group peer leads to superior labour market outcomes, we interpret the latter effect as a result of tighter social connections to group peers. The effect might stem from various sources - referral premiums, a higher job arrival rate, a superior job offer distribution, productivity gains.[26] Table 11 supports the assumption that there is no ex-ante difference between jobs where group peers and cohort peers work.

4.4.3 Results

In this section, we present and discuss our empirical results on whether it is beneficial to use alumni networks to find a new job.

4.4.3.1 Main Results

We estimate the effect of joining a firm with an incumbent group peer vs an incumbent cohort peer using the framework formalized in Eq. 4.4. We consider multiple labour market outcomes. Besides daily wages, we also investigate the effects on job turnover behaviour. If peers provide information about jobs that match skills and/or preferences of workers better, we expect workers who join their group peers to stay at the workplace longer. Moreover, access to better-paying jobs might constitute a part of the benefits that workers get from alumni networks. To further investigate this point, we construct pay ranks of firms and occupations. We check if workers tend to join their group peers at better-paying workplaces/occupations/industries.[27]

As Table 12 suggests, joining a workplace with an incumbent group peer results in 5.8% higher daily wages than joining a workplace with an incumbent cohort peer. The probability of leaving the new employer is 5.9 percentage points lower, which means

[25]As the graphs represent raw means, the difference in levels between the two groups is not informative.

[26]Alternative interpretation that we are not able to rule out here is that workers have higher reservation wages when they accept job offers from a place where someone works whom they know. Such behaviour could be rationalized by a higher weight of socially closer individuals in interpersonal comparisons.

[27]The pay rank is defined as the rank of a firm fixed effect (occupation fixed effect) in a given year from a daily wage regression on the entire Danish working population (1985-2016) with individual fixed effects, year fixed effects and age polynomials. Industry fixed effects are calculated as employment weighted averages of firm fixed effects. The lowest rank is 1 and the highest is 100.

Table 12: The Effect of Working Together: Baseline Regressions

	Log Daily Wage	Firm Change	Occupation Pay Rank	Industry Pay Rank	Firm Pay Rank
Group Peer	0.0578***	-0.0591***	2.836***	3.135***	3.861***
	(0.0150)	(0.0126)	(0.721)	(0.906)	(0.843)
Cohort Peer	0.0791***	-0.0677***	2.806***	5.108***	6.436***
	(0.00667)	(0.00607)	(0.347)	(0.471)	(0.471)
R-squared	0.617	0.146	0.502	0.606	0.605
Observations	170,725	139,998	150,114	143,794	143,794

Notes: This table shows estimates from the model as specified in Eq. 4.4. The standard errors are cluster at the peer group level. Occupation/Firm/Industry Pay Ranks are based on corresponding fixed effects from regressions on the whole Danish working population with individual fixed effects, year fixed effects and age polynomials. *** p<0.01, ** p<0.05, * p<0.1

that job matches formed in a treatment group tend to be significantly more stable. Not all the return to alumni networks to find jobs, therefore, corresponds to within-firm "referral premiums", as joining a group peer is associated with joining better-paying firms, industries and occupations.

4.4.3.2 Heterogeneity

Table 13 investigates heterogeneous returns to joining a group peer in terms of daily wages. We interact the treatment variable with various predetermined characteristics of the worker. The results largely mirror the heterogeneity results from Section 4.3.3.2. Students experience the largest returns of joining a group peer (over joining a cohort peer) at early stages in their careers. There are no statistically significant gender differences. Students with a higher ability (measured by high school GPA) and students from the wealthy background (measured by a father in the top 1% of income distribution in the year before matriculation) further experience more substantial wage effects of joining a "network" job.[28]

4.5 Conclusion

This paper studies how social connections between fellow students shape their career outcomes. Our empirical strategy relies on the peer group randomization in a large business education program in Denmark, many of whose graduates later are part of the country's top earners. Using extensive Danish labour market register data, we follow the individual careers of former academic peers. To identify the effects of peer interactions on career trajectories, we compare propensities to choose similar jobs between pairs of students from the same peer groups and pairs of students from the same cohorts but a different peer group. We find that the professional lives of former group peers tend to be significantly more similar. The effect underlying these pattern is an "excess"

[28]Here stronger effects might include 1) higher returns to social interactions and 2) a higher probability that a given job-to-job transition was caused by some sort of social interaction or 3) both.

Table 13: The Effect of Working Together: Heterogeneous Effects

	(1)	(2)	(3)	(4)
Group Peer	0.0561***	0.0580***	0.0446***	0.0475***
	(0.0159)	(0.0156)	(0.0166)	(0.0161)
Group Peer×YSM		-0.00821***		-0.00830***
		(0.00228)		(0.00235)
Group Peer×Female		-0.00626		-0.0274
		(0.0273)		(0.0287)
Group Peer×High GPA			0.0587**	0.0770***
			(0.0249)	(0.0246)
Group Peer×Father Top 1%			0.0672**	0.0851***
			(0.0293)	(0.0294)
Cohort Peer	0.0773***	0.0769***	0.0782***	0.0776***
	(0.00727)	(0.00729)	(0.00777)	(0.00780)
R-squared	0.636	0.636	0.638	0.638
Observations	159,536	159,536	134,082	134,082

Notes: In all columns log daily wage is a dependent variable. Group Peer variable is interacted with the number of years after graduation, a female dummy, a dummy for high school GPA above cohort median and a dummy for having a father in Top 1% of Danish disposable income distribution year before matriculation. Variables in interaction terms were demeaned. The standard errors are cluster at the peer group level. *** p<0.01, ** p<0.05, * p<0.1

tendency of the former group peers to be employed at the same workplaces. The evidence suggests that ”excess” similarities between group peers are more likely to be driven by contemporaneous alumni interactions than by persistence of influences from the study period. Our analysis reveals that alumni networks persist for years after graduation, although their importance tends to decrease. The effect of social interactions with peers on career choices exhibits strong homophily by gender and origin. Importantly, students of higher ability and who stem from wealthy family background experience the strongest effects of peers on careers.

Moreover, we ask if students following group peers experience significant career improvements. To answer this question, we implement a strategy based on comparing two types of job-to-job transitions - joining a firm with a former peer group peer and joining a firm with a former cohort peer. We show that students that join their group peers benefit from higher wage growth and gain access to more stable and better-paying jobs. Again, these effects are most pronounced for high-ability students coming from affluent families at the earlier stages of their careers.

In this paper, we provide evidence for the significant benefits of alumni networks for students at the start of their careers. Universities who want to maximize their student’s educational returns are therefore well advised to foster alumni interactions through alumni events. Furthermore, our results indicate that fostering alumni networks raises equity concerns.

References

Anelli, M., & Peri, G. (2017). The Effects of High School Peers' Gender on College Major, College Performance and Income. *The Economic Journal*, *129*(618), 553–602.

Angrist, J. (2014). The Perils of Peer Effects. *Labour Economics*, *30*, 98–108.

Bayer, P., Ross, S., & Topa, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy*, *116*(6), 1150–1196.

Bjerge, B., & Skibsted, M. (2016). Do Peers Matter? - Impacts of Peers on Master's Choice and Labour Market Outcomes. *Working Paper*.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2013). Under Pressure? The Effect of Peers on Outcomes of Young Adults. *Journal of Labor Economics*, *31*(1), 119–153.

Brenøe, A., & Zölitz, U. (2020). Exposure to More Female Peers Widens the Gender Gap in STEM Participation. *Journal of Labor Economics, 38*(4).

Brown, M., Setren, E., & Topa, G. (2016). Do Informal Referrals Lead to Better Matches? Evidence from a Firm's Employee Referral System. *Journal of Labor Economics*, *34*(1), 161–209.

Burks, S., Cowgill, B., Hoffman, M., & Housman, M. (2015). The Value of Hiring through Employee Referrals. *The Quarterly Journal of Economics*, *130*(2), 805–839.

Caldwell, S., & Harmon, N. (2018). Outside Options, Bargaining, and Wages: Evidence from Coworker Networks. *Working Paper*.

Carrell, S., Fullerton, R. L., & West, J. (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, *27*(3), 439–464.

Cingano, F., & Rosolia, A. (2012). People I Know: Job Search and Social Networks. *Journal of Labor Economics*, *30*(2), 291–332.

Damm, A. P. (2009). Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence. *Journal of Labor Economics*, *27*(2), 281–314.

Dustmann, C., Glitz, A., Schönberg, U., & Brücker, H. (2016). Referral-based Job Search Networks. *Review of Economic Studies*, *83*(2), 514–546.

Edin, P., Fredriksson, P., & Åslund, O. (2003). Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment. *The Quarterly Journal of Economics*, *118*(1), 329–357.

Eliason, M., Hensvik, L., Kramarz, F., & Skans, N. (2019). *Social Connections and the Sorting of Workers to Firms* (CEPR Discussion Papers No. 13672). C.E.P.R. Discussion Papers.

Fafchamps, M., & Gubert, F. (2007). Risk Sharing and Network Formation. *American Economic Review*, *97*(2), 75–79.

Feld, J., Salamanca, N., & Zölitz, U. (2020). Are Professors Worth It? The Value-Added and Costs of Tutorial Instructors. *Journal of Human Resources*, *55*(3), 836–863.

Feld, J., & Zölitz, U. (2017). Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. *Journal of Labor Economics*, *35*(2), 387–428.

Feld, J., & Zölitz, U. (2018). *Peers from Venus and Mars – Higher-Achieving Men Foster Gender Gaps in Major Choice and Labor Market Outcomes* (Working Paper). In: Cesifo Area Conferences : Economics of Education.

Fischer, A., Gorshkov, A., Sandoy, T. M., & Walldorf, J. (2021). *University Peers and Labour Market Gender Gaps*.

Glitz, A. (2017). Coworker Networks in the Labour Market. *Labour Economics*, *44*, 218–230.

Glitz, A., & Vejlin, R. M. (2019). Learning through Coworker Referrals. *IZA Discussion Papers, No. 12270*.

Hacamo, I., & Kleiner, K. (2017). Competing for Talent: Firms, Managers, and Social Networks. *Working Paper*.

Hellerstein, J. K., McInerney, M., & Neumark, D. (2011). Neighbors and Coworkers: The Importance of Residential Labor Market Networks. *Journal of Labor Economics*, *29*(4), 659–695.

Hoxby, C. (2000). *Peer Effects in the Classroom: Learning from Gender and Race Variation* (NBER Working Papers).

Jones, T., & Kofoed, M. (2020). Do Peers Influence Occupational Preferences? Evidence from Randomly Assigned Peer Groups at West Point. *Journal of Public Economics*, *184*, 104–154.

Kline, P., & Santos, A. (2012). A Score Based Approach to Wild Bootstrap Inference. *Journal of Econometric Methods*, *1*(1), 23–41.

Kramarz, F., & Skans, O. (2014). When Strong Ties are Strong: Networks and Youth Labour Market Entry. *Review of Economic Studies*, *81*(3), 1164–1200.

Lerner, J., & Malmendier, U. (2013). With a Little Help from my (Random) Friends: Success and Failure in Post-Business School Entrepreneurship. *The Review of Financial Studies*, *26*(10), 2411–2452.

Lyle, D. S. (2007). Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point. *The Review of Economics and Statistics*, *89*(2), 289–299.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, *60*(3), 531–542.

Sacerdote, B. (2014). Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward? *Annual Review of Economics*, *6*(1), 253–272.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, *116*(2), 681–704.

Saygin, P. O., Weber, A., & Weynandt, M. A. (2019). Coworkers, Networks, and Job-Search Outcomes among Displaced Workers. *ILR Review*.

Schmutte, I. (2015). Job Referral Networks and the Determination of Earnings in Local Labor Markets. *Journal of Labor Economics*, *33*(1), 1–32.

Shue, K. (2013). Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers. *Review of Financial Studies*, *26*(6), 1401–1442.

University of Chicago Booth. (2018). The Value of Booth Network [Accessed: 2018-10-30].

Zhu, M. (2018). Networks through College Classmates: Effects of Referrals for Men and Women. *Working Paper*.

Zimmerman, D. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *The Review of Economics and Statistics*, *85*(1), 9–23.

Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review*, *109*(1), 1–47.

TITLER I PH.D.SERIEN:

2004

1. Martin Grieger
Internet-based Electronic Marketplaces and Supply Chain Management

2. Thomas Basbøll
LIKENESS
A Philosophical Investigation

3. Morten Knudsen
Beslutningens vaklen
En systemteoretisk analyse of moderniseringen af et amtskommunalt sundhedsvæsen 1980-2000

4. Lars Bo Jeppesen
Organizing Consumer Innovation
A product development strategy that is based on online communities and allows some firms to benefit from a distributed process of innovation by consumers

5. Barbara Dragsted
SEGMENTATION IN TRANSLATION AND TRANSLATION MEMORY SYSTEMS
An empirical investigation of cognitive segmentation and effects of integrating a TM system into the translation process

6. Jeanet Hardis
Sociale partnerskaber
Et socialkonstruktivistisk casestudie af partnerskabsaktørers virkelighedsopfattelse mellem identitet og legitimitet

7. Henriette Hallberg Thygesen
System Dynamics in Action

8. Carsten Mejer Plath
Strategisk Økonomistyring

9. Annemette Kjærgaard
Knowledge Management as Internal Corporate Venturing
– a Field Study of the Rise and Fall of a Bottom-Up Process

10. Knut Arne Hovdal
De profesjonelle i endring
Norsk ph.d., ej til salg gennem Samfundslitteratur

11. Søren Jeppesen
Environmental Practices and Greening Strategies in Small Manufacturing Enterprises in South Africa
– A Critical Realist Approach

12. Lars Frode Frederiksen
Industriel forskningsledelse
– på sporet af mønstre og samarbejde i danske forskningsintensive virksomheder

13. Martin Jes Iversen
The Governance of GN Great Nordic
– in an age of strategic and structural transitions 1939-1988

14. Lars Pynt Andersen
The Rhetorical Strategies of Danish TV Advertising
A study of the first fifteen years with special emphasis on genre and irony

15. Jakob Rasmussen
Business Perspectives on E-learning

16. Sof Thrane
The Social and Economic Dynamics of Networks
– a Weberian Analysis of Three Formalised Horizontal Networks

17. Lene Nielsen
Engaging Personas and Narrative Scenarios – a study on how a user-centered approach influenced the perception of the design process in the e-business group at AstraZeneca

18. S.J Valstad
Organisationsidentitet
Norsk ph.d., ej til salg gennem Samfundslitteratur

19. Thomas Lyse Hansen
Six Essays on Pricing and Weather risk in Energy Markets

20. Sabine Madsen
Emerging Methods – An Interpretive Study of ISD Methods in Practice

21. Evis Sinani
The Impact of Foreign Direct Investment on Efficiency, Productivity Growth and Trade: An Empirical Investigation

22. Bent Meier Sørensen
Making Events Work Or, How to Multiply Your Crisis

23. Pernille Schnoor
Brand Ethos
Om troværdige brand- og virksomhedsidentiteter i et retorisk og diskursteoretisk perspektiv

24. Sidsel Fabech
Von welchem Österreich ist hier die Rede?
Diskursive forhandlinger og magtkampe mellem rivaliserende nationale identitetskonstruktioner i østrigske pressediskurser

25. Klavs Odgaard Christensen
Sprogpolitik og identitetsdannelse i flersprogede forbundsstater
Et komparativt studie af Schweiz og Canada

26. Dana B. Minbaeva
Human Resource Practices and Knowledge Transfer in Multinational Corporations

27. Holger Højlund
Markedets politiske fornuft
Et studie af velfærdens organisering i perioden 1990-2003

28. Christine Mølgaard Frandsen
A.s erfaring
Om mellemværendets praktik i en transformation af mennesket og subjektiviteten

29. Sine Nørholm Just
The Constitution of Meaning – A Meaningful Constitution?
Legitimacy, identity, and public opinion in the debate on the future of Europe

2005

1. Claus J. Varnes
Managing product innovation through rules – The role of formal and structured methods in product development

2. Helle Hedegaard Hein
Mellem konflikt og konsensus – Dialogudvikling på hospitalsklinikker

3. Axel Rosenø
Customer Value Driven Product Innovation – A Study of Market Learning in New Product Development

4. Søren Buhl Pedersen
Making space
An outline of place branding

5. Camilla Funck Ellehave
Differences that Matter
An analysis of practices of gender and organizing in contemporary workplaces

6. Rigmor Madeleine Lond
Styring af kommunale forvaltninger

7. Mette Aagaard Andreassen
Supply Chain versus Supply Chain Benchmarking as a Means to Managing Supply Chains

8. Caroline Aggestam-Pontoppidan
From an idea to a standard
The UN and the global governance of accountants' competence

9. Norsk ph.d.

10. Vivienne Heng Ker-ni
An Experimental Field Study on the

Effectiveness of Grocer Media Advertising
Measuring Ad Recall and Recognition, Purchase Intentions and Short-Term Sales

11. Allan Mortensen
Essays on the Pricing of Corporate Bonds and Credit Derivatives

12. Remo Stefano Chiari
Figure che fanno conoscere
Itinerario sull'idea del valore cognitivo e espressivo della metafora e di altri tropi da Aristotele e da Vico fino al cognitivismo contemporaneo

13. Anders McIlquham-Schmidt
Strategic Planning and Corporate Performance
An integrative research review and a meta-analysis of the strategic planning and corporate performance literature from 1956 to 2003

14. Jens Geersbro
The TDF – PMI Case
Making Sense of the Dynamics of Business Relationships and Networks

15 Mette Andersen
Corporate Social Responsibility in Global Supply Chains
Understanding the uniqueness of firm behaviour

16. Eva Boxenbaum
Institutional Genesis: Micro – Dynamic Foundations of Institutional Change

17. Peter Lund-Thomsen
Capacity Development, Environmental Justice NGOs, and Governance: The Case of South Africa

18. Signe Jarlov
Konstruktioner af offentlig ledelse

19. Lars Stæhr Jensen
Vocabulary Knowledge and Listening Comprehension in English as a Foreign Language
An empirical study employing data elicited from Danish EFL learners

20. Christian Nielsen
Essays on Business Reporting
Production and consumption of strategic information in the market for information

21. Marianne Thejls Fischer
Egos and Ethics of Management Consultants

22. Annie Bekke Kjær
Performance management i Proces-innovation
– belyst i et social-konstruktivistisk perspektiv

23. Suzanne Dee Pedersen
GENTAGELSENS METAMORFOSE
Om organisering af den kreative gøren i den kunstneriske arbejdspraksis

24. Benedikte Dorte Rosenbrink
Revenue Management
Økonomiske, konkurrencemæssige & organisatoriske konsekvenser

25. Thomas Riise Johansen
Written Accounts and Verbal Accounts
The Danish Case of Accounting and Accountability to Employees

26. Ann Fogelgren-Pedersen
The Mobile Internet: Pioneering Users' Adoption Decisions

27. Birgitte Rasmussen
Ledelse i fællesskab – de tillidsvalgtes fornyende rolle

28. Gitte Thit Nielsen
Remerger
– skabende ledelseskræfter i fusion og opkøb

29. Carmine Gioia
A MICROECONOMETRIC ANALYSIS OF MERGERS AND ACQUISITIONS

30. Ole Hinz
Den effektive forandringsleder: pilot, pædagog eller politiker?
Et studie i arbejdslederes meningstilskrivninger i forbindelse med vellykket gennemførelse af ledelsesinitierede forandringsprojekter

31. Kjell-Åge Gotvassli
Et praksisbasert perspektiv på dynamiske
læringsnettverk i toppidretten
Norsk ph.d., ej til salg gennem Samfundslitteratur

32. Henriette Langstrup Nielsen
Linking Healthcare
An inquiry into the changing performances of web-based technology for asthma monitoring

33. Karin Tweddell Levinsen
Virtuel Uddannelsespraksis
Master i IKT og Læring – et casestudie i hvordan proaktiv proceshåndtering kan forbedre praksis i virtuelle læringsmiljøer

34. Anika Liversage
Finding a Path
Labour Market Life Stories of Immigrant Professionals

35. Kasper Elmquist Jørgensen
Studier i samspillet mellem stat og erhvervsliv i Danmark under 1. verdenskrig

36. Finn Janning
A DIFFERENT STORY
Seduction, Conquest and Discovery

37. Patricia Ann Plackett
Strategic Management of the Radical Innovation Process
Leveraging Social Capital for Market Uncertainty Management

2006

1. Christian Vintergaard
Early Phases of Corporate Venturing

2. Niels Rom-Poulsen
Essays in Computational Finance

3. Tina Brandt Husman
Organisational Capabilities, Competitive Advantage & Project-Based Organisations
The Case of Advertising and Creative Good Production

4. Mette Rosenkrands Johansen
Practice at the top
– how top managers mobilise and use non-financial performance measures

5. Eva Parum
Corporate governance som strategisk kommunikations- og ledelsesværktøj

6. Susan Aagaard Petersen
Culture's Influence on Performance Management: The Case of a Danish Company in China

7. Thomas Nicolai Pedersen
The Discursive Constitution of Organizational Governance – Between unity and differentiation
The Case of the governance of environmental risks by World Bank environmental staff

8. Cynthia Selin
Volatile Visions: Transactons in Anticipatory Knowledge

9. Jesper Banghøj
Financial Accounting Information and Compensation in Danish Companies

10. Mikkel Lucas Overby
Strategic Alliances in Emerging High-Tech Markets: What's the Difference and does it Matter?

11. Tine Aage
External Information Acquisition of Industrial Districts and the Impact of Different Knowledge Creation Dimensions

A case study of the Fashion and Design Branch of the Industrial District of Montebelluna, NE Italy

12. Mikkel Flyverbom
Making the Global Information Society Governable
On the Governmentality of Multi-Stakeholder Networks

13. Anette Grønning
Personen bag
Tilstedevær i e-mail som inter-aktionsform mellem kunde og med-arbejder i dansk forsikringskontekst

14. Jørn Helder
One Company – One Language?
The NN-case

15. Lars Bjerregaard Mikkelsen
Differing perceptions of customer value
Development and application of a tool for mapping perceptions of customer value at both ends of customer-supplier dyads in industrial markets

16. Lise Granerud
Exploring Learning
Technological learning within small manufacturers in South Africa

17. Esben Rahbek Pedersen
Between Hopes and Realities: Reflections on the Promises and Practices of Corporate Social Responsibility (CSR)

18. Ramona Samson
The Cultural Integration Model and European Transformation.
The Case of Romania

2007

1. Jakob Vestergaard
Discipline in The Global Economy
Panopticism and the Post-Washington Consensus

2. Heidi Lund Hansen
Spaces for learning and working
A qualitative study of change of work, management, vehicles of power and social practices in open offices

3. Sudhanshu Rai
Exploring the internal dynamics of software development teams during user analysis
A tension enabled Institutionalization Model; "Where process becomes the objective"

4. Norsk ph.d.
Ej til salg gennem Samfundslitteratur

5. Serden Ozcan
EXPLORING HETEROGENEITY IN ORGANIZATIONAL ACTIONS AND OUTCOMES
A Behavioural Perspective

6. Kim Sundtoft Hald
Inter-organizational Performance Measurement and Management in Action
– An Ethnography on the Construction of Management, Identity and Relationships

7. Tobias Lindeberg
Evaluative Technologies
Quality and the Multiplicity of Performance

8. Merete Wedell-Wedellsborg
Den globale soldat
Identitetsdannelse og identitetsledelse i multinationale militære organisationer

9. Lars Frederiksen
Open Innovation Business Models
Innovation in firm-hosted online user communities and inter-firm project ventures in the music industry
– A collection of essays

10. Jonas Gabrielsen
Retorisk toposlære – fra statisk 'sted' til persuasiv aktivitet

11. Christian Moldt-Jørgensen
Fra meningsløs til meningsfuld evaluering.
Anvendelsen af studentertilfredshedsmålinger på de korte og mellemlange videregående uddannelser set fra et psykodynamisk systemperspektiv

12. Ping Gao
Extending the application of actor-network theory
Cases of innovation in the telecommunications industry

13. Peter Mejlby
Frihed og fængsel, en del af den samme drøm?
Et phronetisk baseret casestudie af frigørelsens og kontrollens sameksistens i værdibaseret ledelse!

14. Kristina Birch
Statistical Modelling in Marketing

15. Signe Poulsen
Sense and sensibility:
The language of emotional appeals in insurance marketing

16. Anders Bjerre Trolle
Essays on derivatives pricing and dynamic asset allocation

17. Peter Feldhütter
Empirical Studies of Bond and Credit Markets

18. Jens Henrik Eggert Christensen
Default and Recovery Risk Modeling and Estimation

19. Maria Theresa Larsen
Academic Enterprise: A New Mission for Universities or a Contradiction in Terms?
Four papers on the long-term implications of increasing industry involvement and commercialization in academia

20. Morten Wellendorf
Postimplementering af teknologi i den offentlige forvaltning
Analyser af en organisations kontinuerlige arbejde med informationsteknologi

21. Ekaterina Mhaanna
Concept Relations for Terminological Process Analysis

22. Stefan Ring Thorbjørnsen
Forsvaret i forandring
Et studie i officerers kapabiliteter under påvirkning af omverdenens forandringspres mod øget styring og læring

23. Christa Breum Amhøj
Det selvskabte medlemskab om managementstaten, dens styringsteknologier og indbyggere

24. Karoline Bromose
Between Technological Turbulence and Operational Stability
– An empirical case study of corporate venturing in TDC

25. Susanne Justesen
Navigating the Paradoxes of Diversity in Innovation Practice
– A Longitudinal study of six very different innovation processes – in practice

26. Luise Noring Henler
Conceptualising successful supply chain partnerships
– Viewing supply chain partnerships from an organisational culture perspective

27. Mark Mau
Kampen om telefonen
Det danske telefonvæsen under den tyske besættelse 1940-45

28. Jakob Halskov
The semiautomatic expansion of existing terminological ontologies using knowledge patterns discovered

on the WWW – an implementation and evaluation

29. Gergana Koleva
European Policy Instruments Beyond Networks and Structure: The Innovative Medicines Initiative

30. Christian Geisler Asmussen
Global Strategy and International Diversity: A Double-Edged Sword?

31. Christina Holm-Petersen
Stolthed og fordom
Kultur- og identitetsarbejde ved skabelsen af en ny sengeafdeling gennem fusion

32. Hans Peter Olsen
Hybrid Governance of Standardized States
Causes and Contours of the Global Regulation of Government Auditing

33. Lars Bøge Sørensen
Risk Management in the Supply Chain

34. Peter Aagaard
Det unikkes dynamikker
De institutionelle mulighedsbetingelser bag den individuelle udforskning i professionelt og frivilligt arbejde

35. Yun Mi Antorini
Brand Community Innovation
An Intrinsic Case Study of the Adult Fans of LEGO Community

36. Joachim Lynggaard Boll
Labor Related Corporate Social Performance in Denmark
Organizational and Institutional Perspectives

2008

1. Frederik Christian Vinten
Essays on Private Equity

2. Jesper Clement
Visual Influence of Packaging Design on In-Store Buying Decisions

3. Marius Brostrøm Kousgaard
Tid til kvalitetsmåling?
– Studier af indrulleringsprocesser i forbindelse med introduktionen af kliniske kvalitetsdatabaser i speciallægepraksissektoren

4. Irene Skovgaard Smith
Management Consulting in Action
Value creation and ambiguity in client-consultant relations

5. Anders Rom
Management accounting and integrated information systems
How to exploit the potential for management accounting of information technology

6. Marina Candi
Aesthetic Design as an Element of Service Innovation in New Technology-based Firms

7. Morten Schnack
Teknologi og tværfaglighed
– en analyse af diskussionen omkring indførelse af EPJ på en hospitalsafdeling

8. Helene Balslev Clausen
Juntos pero no revueltos – un estudio sobre emigrantes norteamericanos en un pueblo mexicano

9. Lise Justesen
Kunsten at skrive revisionsrapporter.
En beretning om forvaltningsrevisionens beretninger

10. Michael E. Hansen
The politics of corporate responsibility: CSR and the governance of child labor and core labor rights in the 1990s

11. Anne Roepstorff
Holdning for handling – en etnologisk undersøgelse af Virksomheders Sociale Ansvar/CSR

12. Claus Bajlum
Essays on Credit Risk and Credit Derivatives

13. Anders Bojesen
The Performative Power of Competence – an Inquiry into Subjectivity and Social Technologies at Work

14. Satu Reijonen
Green and Fragile
A Study on Markets and the Natural Environment

15. Ilduara Busta
Corporate Governance in Banking
A European Study

16. Kristian Anders Hvass
A Boolean Analysis Predicting Industry Change: Innovation, Imitation & Business Models
The Winning Hybrid: A case study of isomorphism in the airline industry

17. Trine Paludan
De uvidende og de udviklingsparate
Identitet som mulighed og restriktion blandt fabriksarbejdere på det aftayloriserede fabriksgulv

18. Kristian Jakobsen
Foreign market entry in transition economies: Entry timing and mode choice

19. Jakob Elming
Syntactic reordering in statistical machine translation

20. Lars Brømsøe Termansen
Regional Computable General Equilibrium Models for Denmark
Three papers laying the foundation for regional CGE models with agglomeration characteristics

21. Mia Reinholt
The Motivational Foundations of Knowledge Sharing

22. Frederikke Krogh-Meibom
The Co-Evolution of Institutions and Technology
– A Neo-Institutional Understanding of Change Processes within the Business Press – the Case Study of Financial Times

23. Peter D. Ørberg Jensen
OFFSHORING OF ADVANCED AND HIGH-VALUE TECHNICAL SERVICES: ANTECEDENTS, PROCESS DYNAMICS AND FIRMLEVEL IMPACTS

24. Pham Thi Song Hanh
Functional Upgrading, Relational Capability and Export Performance of Vietnamese Wood Furniture Producers

25. Mads Vangkilde
Why wait?
An Exploration of first-mover advantages among Danish e-grocers through a resource perspective

26. Hubert Buch-Hansen
Rethinking the History of European Level Merger Control
A Critical Political Economy Perspective

2009

1. Vivian Lindhardsen
From Independent Ratings to Communal Ratings: A Study of CWA Raters' Decision-Making Behaviours

2. Guðrið Weihe
Public-Private Partnerships: Meaning and Practice

3. Chris Nøkkentved
Enabling Supply Networks with Collaborative Information Infrastructures
An Empirical Investigation of Business Model Innovation in Supplier Relationship Management

4. Sara Louise Muhr
Wound, Interrupted – On the Vulnerability of Diversity Management

5. Christine Sestoft
Forbrugeradfærd i et Stats- og Livsformsteoretisk perspektiv

6. Michael Pedersen
Tune in, Breakdown, and Reboot: On the production of the stress-fit self-managing employee

7. Salla Lutz
Position and Reposition in Networks – Exemplified by the Transformation of the Danish Pine Furniture Manufacturers

8. Jens Forssbæck
Essays on market discipline in commercial and central banking

9. Tine Murphy
Sense from Silence – A Basis for Organised Action
How do Sensemaking Processes with Minimal Sharing Relate to the Reproduction of Organised Action?

10. Sara Malou Strandvad
Inspirations for a new sociology of art: A sociomaterial study of development processes in the Danish film industry

11. Nicolaas Mouton
On the evolution of social scientific metaphors:
A cognitive-historical enquiry into the divergent trajectories of the idea that collective entities – states and societies, cities and corporations – are biological organisms.

12. Lars Andreas Knutsen
Mobile Data Services:
Shaping of user engagements

13. Nikolaos Theodoros Korfiatis
Information Exchange and Behavior
A Multi-method Inquiry on Online Communities

14. Jens Albæk
Forestillinger om kvalitet og tværfaglighed på sygehuse
– skabelse af forestillinger i læge- og plejegrupperne angående relevans af nye idéer om kvalitetsudvikling gennem tolkningsprocesser

15. Maja Lotz
The Business of Co-Creation – and the Co-Creation of Business

16. Gitte P. Jakobsen
Narrative Construction of Leader Identity in a Leader Development Program Context

17. Dorte Hermansen
"Living the brand" som en brandorienteret dialogisk praxis:
Om udvikling af medarbejdernes brandorienterede dømmekraft

18. Aseem Kinra
Supply Chain (logistics) Environmental Complexity

19. Michael Nørager
How to manage SMEs through the transformation from non innovative to innovative?

20. Kristin Wallevik
Corporate Governance in Family Firms
The Norwegian Maritime Sector

21. Bo Hansen Hansen
Beyond the Process
Enriching Software Process Improvement with Knowledge Management

22. Annemette Skot-Hansen
Franske adjektivisk afledte adverbier, der tager præpositionssyntagmer indledt med præpositionen à som argumenter
En valensgrammatisk undersøgelse

23. Line Gry Knudsen
Collaborative R&D Capabilities
In Search of Micro-Foundations

24. Christian Scheuer
Employers meet employees
Essays on sorting and globalization

25. Rasmus Johnsen
The Great Health of Melancholy
A Study of the Pathologies of Performativity

26. Ha Thi Van Pham
Internationalization, Competitiveness Enhancement and Export Performance of Emerging Market Firms:
Evidence from Vietnam

27. Henriette Balieu
Kontrolbegrebets betydning for kausativalternationen i spansk
En kognitiv-typologisk analyse

2010

1. Yen Tran
Organizing Innovationin Turbulent Fashion Market
Four papers on how fashion firms create and appropriate innovation value

2. Anders Raastrup Kristensen
Metaphysical Labour
Flexibility, Performance and Commitment in Work-Life Management

3. Margrét Sigrún Sigurdardottir
Dependently independent
Co-existence of institutional logics in the recorded music industry

4. Ásta Dis Óladóttir
Internationalization from a small domestic base:
An empirical analysis of Economics and Management

5. Christine Secher
E-deltagelse i praksis – politikernes og forvaltningens medkonstruktion og konsekvenserne heraf

6. Marianne Stang Våland
What we talk about when we talk about space:
End User Participation between Processes of Organizational and Architectural Design

7. Rex Degnegaard
Strategic Change Management
Change Management Challenges in the Danish Police Reform

8. Ulrik Schultz Brix
Værdi i rekruttering – den sikre beslutning
En pragmatisk analyse af perception og synliggørelse af værdi i rekrutterings- og udvælgelsesarbejdet

9. Jan Ole Similä
Kontraktsledelse
Relasjonen mellom virksomhetsledelse og kontraktshåndtering, belyst via fire norske virksomheter

10. Susanne Boch Waldorff
Emerging Organizations: In between local translation, institutional logics and discourse

11. Brian Kane
Performance Talk
Next Generation Management of Organizational Performance

12. Lars Ohnemus
Brand Thrust: Strategic Branding and Shareholder Value
An Empirical Reconciliation of two Critical Concepts

13. Jesper Schlamovitz
Håndtering af usikkerhed i film- og byggeprojekter

14. Tommy Moesby-Jensen
Det faktiske livs forbindtlighed
Førsokratisk informeret, ny-aristotelisk ἦθος-tænkning hos Martin Heidegger

15. Christian Fich
Two Nations Divided by Common Values
French National Habitus and the Rejection of American Power

16. Peter Beyer
Processer, sammenhængskraft og fleksibilitet
Et empirisk casestudie af omstillingsforløb i fire virksomheder

17. Adam Buchhorn
Markets of Good Intentions
Constructing and Organizing Biogas Markets Amid Fragility and Controversy

18. Cecilie K. Moesby-Jensen
Social læring og fælles praksis
Et mixed method studie, der belyser læringskonsekvenser af et lederkursus for et praksisfællesskab af offentlige mellemledere

19. Heidi Boye
Fødevarer og sundhed i senmodernismen
– En indsigt i hyggefænomenet og de relaterede fødevarepraksisser

20. Kristine Munkgård Pedersen
Flygtige forbindelser og midlertidige mobiliseringer
Om kulturel produktion på Roskilde Festival

21. Oliver Jacob Weber
Causes of Intercompany Harmony in Business Markets – An Empirical Investigation from a Dyad Perspective

22. Susanne Ekman
Authority and Autonomy
Paradoxes of Modern Knowledge Work

23. Anette Frey Larsen
Kvalitetsledelse på danske hospitaler
– Ledelsernes indflydelse på introduktion og vedligeholdelse af kvalitetsstrategier i det danske sundhedsvæsen

24. Toyoko Sato
Performativity and Discourse: Japanese Advertisements on the Aesthetic Education of Desire

25. Kenneth Brinch Jensen
Identifying the Last Planner System
Lean management in the construction industry

26. Javier Busquets
Orchestrating Network Behavior for Innovation

27. Luke Patey
The Power of Resistance: India's National Oil Company and International Activism in Sudan

28. Mette Vedel
Value Creation in Triadic Business Relationships. Interaction, Interconnection and Position

29. Kristian Tørning
Knowledge Management Systems in Practice – A Work Place Study

30. Qingxin Shi
An Empirical Study of Thinking Aloud Usability Testing from a Cultural Perspective

31. Tanja Juul Christiansen
Corporate blogging: Medarbejderes kommunikative handlekraft

32. Malgorzata Ciesielska
Hybrid Organisations.
A study of the Open Source – business setting

33. Jens Dick-Nielsen
Three Essays on Corporate Bond Market Liquidity

34. Sabrina Speiermann
Modstandens Politik
Kampagnestyring i Velfærdsstaten.
En diskussion af trafikkampagners styringspotentiale

35. Julie Uldam
Fickle Commitment. Fostering political engagement in 'the flighty world of online activism'

36. Annegrete Juul Nielsen
Traveling technologies and transformations in health care

37. Athur Mühlen-Schulte
Organising Development
Power and Organisational Reform in the United Nations Development Programme

38. Louise Rygaard Jonas
Branding på butiksgulvet
Et case-studie af kultur- og identitets-arbejdet i Kvickly

2011

1. Stefan Fraenkel
Key Success Factors for Sales Force Readiness during New Product Launch
A Study of Product Launches in the Swedish Pharmaceutical Industry

2. Christian Plesner Rossing
International Transfer Pricing in Theory and Practice

3. Tobias Dam Hede
Samtalekunst og ledelsesdisciplin
– en analyse af coachingsdiskursens genealogi og governmentality

4. Kim Pettersson
Essays on Audit Quality, Auditor Choice, and Equity Valuation

5. Henrik Merkelsen
The expert-lay controversy in risk research and management. Effects of institutional distances. Studies of risk definitions, perceptions, management and communication

6. Simon S. Torp
Employee Stock Ownership:
Effect on Strategic Management and Performance

7. Mie Harder
Internal Antecedents of Management Innovation

8. Ole Helby Petersen
Public-Private Partnerships: Policy and Regulation – With Comparative and Multi-level Case Studies from Denmark and Ireland

9. Morten Krogh Petersen
'Good' Outcomes. Handling Multiplicity in Government Communication

10. Kristian Tangsgaard Hvelplund
Allocation of cognitive resources in translation - an eye-tracking and key-logging study

11. Moshe Yonatany
The Internationalization Process of Digital Service Providers

12. Anne Vestergaard
Distance and Suffering
Humanitarian Discourse in the age of Mediatization

13. Thorsten Mikkelsen
Personligsheds indflydelse på forretningsrelationer

14. Jane Thostrup Jagd
Hvorfor fortsætter fusionsbølgen udover "the tipping point"?
– en empirisk analyse af information og kognitioner om fusioner

15. Gregory Gimpel
Value-driven Adoption and Consumption of Technology: Understanding Technology Decision Making

16. Thomas Stengade Sønderskov
Den nye mulighed
Social innovation i en forretningsmæssig kontekst

17. Jeppe Christoffersen
Donor supported strategic alliances in developing countries

18. Vibeke Vad Baunsgaard
Dominant Ideological Modes of Rationality: Cross functional

integration in the process of product innovation

19. Throstur Olaf Sigurjonsson
Governance Failure and Icelands's Financial Collapse

20. Allan Sall Tang Andersen
Essays on the modeling of risks in interest-rate and inflation markets

21. Heidi Tscherning
Mobile Devices in Social Contexts

22. Birgitte Gorm Hansen
Adapting in the Knowledge Economy
Lateral Strategies for Scientists and Those Who Study Them

23. Kristina Vaarst Andersen
Optimal Levels of Embeddedness
The Contingent Value of Networked Collaboration

24. Justine Grønbæk Pors
Noisy Management
A History of Danish School Governing from 1970-2010

25. Stefan Linder
Micro-foundations of Strategic Entrepreneurship
Essays on Autonomous Strategic Action

26. Xin Li
Toward an Integrative Framework of National Competitiveness
An application to China

27. Rune Thorbjørn Clausen
Værdifuld arkitektur
Et eksplorativt studie af bygningers rolle i virksomheders værdiskabelse

28. Monica Viken
Markedsundersøkelser som bevis i varemerke- og markedsføringsrett

29. Christian Wymann
Tattooing
The Economic and Artistic Constitution of a Social Phenomenon

30. Sanne Frandsen
Productive Incoherence
A Case Study of Branding and Identity Struggles in a Low-Prestige Organization

31. Mads Stenbo Nielsen
Essays on Correlation Modelling

32. Ivan Häuser
Følelse og sprog
Etablering af en ekspressiv kategori, eksemplificeret på russisk

33. Sebastian Schwenen
Security of Supply in Electricity Markets

2012

1. Peter Holm Andreasen
The Dynamics of Procurement Management
- A Complexity Approach

2. Martin Haulrich
Data-Driven Bitext Dependency Parsing and Alignment

3. Line Kirkegaard
Konsulenten i den anden nat
En undersøgelse af det intense arbejdsliv

4. Tonny Stenheim
Decision usefulness of goodwill under IFRS

5. Morten Lind Larsen
Produktivitet, vækst og velfærd
Industrirådet og efterkrigstidens Danmark 1945 - 1958

6. Petter Berg
Cartel Damages and Cost Asymmetries

7. Lynn Kahle
Experiential Discourse in Marketing
A methodical inquiry into practice and theory

8. Anne Roelsgaard Obling
Management of Emotions in Accelerated Medical Relationships

9. Thomas Frandsen
Managing Modularity of Service Processes Architecture

10. Carina Christine Skovmøller
CSR som noget særligt
Et casestudie om styring og meningsskabelse i relation til CSR ud fra en intern optik

11. Michael Tell
Fradragsbeskæring af selskabers finansieringsudgifter
En skatteretlig analyse af SEL §§ 11, 11B og 11C

12. Morten Holm
Customer Profitability Measurement Models
Their Merits and Sophistication across Contexts

13. Katja Joo Dyppel
Beskatning af derivater
En analyse af dansk skatteret

14. Esben Anton Schultz
Essays in Labor Economics
Evidence from Danish Micro Data

15. Carina Risvig Hansen
"Contracts not covered, or not fully covered, by the Public Sector Directive"

16. Anja Svejgaard Pors
Iværksættelse af kommunikation
- patientfigurer i hospitalets strategiske kommunikation

17. Frans Bévort
Making sense of management with logics
An ethnographic study of accountants who become managers

18. René Kallestrup
The Dynamics of Bank and Sovereign Credit Risk

19. Brett Crawford
Revisiting the Phenomenon of Interests in Organizational Institutionalism
The Case of U.S. Chambers of Commerce

20. Mario Daniele Amore
Essays on Empirical Corporate Finance

21. Arne Stjernholm Madsen
The evolution of innovation strategy
Studied in the context of medical device activities at the pharmaceutical company Novo Nordisk A/S in the period 1980-2008

22. Jacob Holm Hansen
Is Social Integration Necessary for Corporate Branding?
A study of corporate branding strategies at Novo Nordisk

23. Stuart Webber
Corporate Profit Shifting and the Multinational Enterprise

24. Helene Ratner
Promises of Reflexivity
Managing and Researching Inclusive Schools

25. Therese Strand
The Owners and the Power: Insights from Annual General Meetings

26. Robert Gavin Strand
In Praise of Corporate Social Responsibility Bureaucracy

27. Nina Sormunen
Auditor's going-concern reporting
Reporting decision and content of the report

28. John Bang Mathiasen
Learning within a product development working practice:
- an understanding anchored in pragmatism

29. Philip Holst Riis
Understanding Role-Oriented Enterprise Systems: From Vendors to Customers

30. Marie Lisa Dacanay
Social Enterprises and the Poor
Enhancing Social Entrepreneurship and Stakeholder Theory

31. Fumiko Kano Glückstad
Bridging Remote Cultures: Cross-lingual concept mapping based on the information receiver's prior-knowledge

32. Henrik Barslund Fosse
Empirical Essays in International Trade

33. Peter Alexander Albrecht
Foundational hybridity and its reproduction
Security sector reform in Sierra Leone

34. Maja Rosenstock
CSR - hvor svært kan det være?
Kulturanalytisk casestudie om udfordringer og dilemmaer med at forankre Coops CSR-strategi

35. Jeanette Rasmussen
Tweens, medier og forbrug
Et studie af 10-12 årige danske børns brug af internettet, opfattelse og forståelse af markedsføring og forbrug

36. Ib Tunby Gulbrandsen
'This page is not intended for a US Audience'
A five-act spectacle on online communication, collaboration & organization.

37. Kasper Aalling Teilmann
Interactive Approaches to Rural Development

38. Mette Mogensen
The Organization(s) of Well-being and Productivity
(Re)assembling work in the Danish Post

39. Søren Friis Møller
From Disinterestedness to Engagement
Towards Relational Leadership In the Cultural Sector

40. Nico Peter Berhausen
Management Control, Innovation and Strategic Objectives – Interactions and Convergence in Product Development Networks

41. Balder Onarheim
Creativity under Constraints
Creativity as Balancing 'Constrainedness'

42. Haoyong Zhou
Essays on Family Firms

43. Elisabeth Naima Mikkelsen
Making sense of organisational conflict
An empirical study of enacted sense-making in everyday conflict at work

2013

1. Jacob Lyngsie
Entrepreneurship in an Organizational Context

2. Signe Groth-Brodersen
Fra ledelse til selvet
En socialpsykologisk analyse af forholdet imellem selvledelse, ledelse og stress i det moderne arbejdsliv

3. Nis Høyrup Christensen
Shaping Markets: A Neoinstitutional Analysis of the Emerging Organizational Field of Renewable Energy in China

4. Christian Edelvold Berg
As a matter of size
THE IMPORTANCE OF CRITICAL MASS AND THE CONSEQUENCES OF SCARCITY FOR TELEVISION MARKETS

5. Christine D. Isakson
Coworker Influence and Labor Mobility
Essays on Turnover, Entrepreneurship and Location Choice in the Danish Maritime Industry

6. Niels Joseph Jerne Lennon
Accounting Qualities in Practice
Rhizomatic stories of representational faithfulness, decision making and control

7. Shannon O'Donnell
Making Ensemble Possible
How special groups organize for collaborative creativity in conditions of spatial variability and distance

8. Robert W. D. Veitch
Access Decisions in a
Partly-Digital World
Comparing Digital Piracy and Legal
Modes for Film and Music

9. Marie Mathiesen
Making Strategy Work
An Organizational Ethnography

10. Arisa Shollo
The role of business intelligence in
organizational decision-making

11. Mia Kaspersen
The construction of social and
environmental reporting

12. Marcus Møller Larsen
The organizational design of offshoring

13. Mette Ohm Rørdam
EU Law on Food Naming
The prohibition against misleading
names in an internal market context

14. Hans Peter Rasmussen
GIV EN GED!
Kan giver-idealtyper forklare støtte
til velgørenhed og understøtte
relationsopbygning?

15. Ruben Schachtenhaufen
Fonetisk reduktion i dansk

16. Peter Koerver Schmidt
Dansk CFC-beskatning
I et internationalt og komparativt
perspektiv

17. Morten Froholdt
Strategi i den offentlige sektor
En kortlægning af styringsmæssig
kontekst, strategisk tilgang, samt
anvendte redskaber og teknologier for
udvalgte danske statslige styrelser

18. Annette Camilla Sjørup
Cognitive effort in metaphor translation
An eye-tracking and key-logging study

19. Tamara Stucchi
The Internationalization
of Emerging Market Firms:
A Context-Specific Study

20. Thomas Lopdrup-Hjorth
"Let's Go Outside":
The Value of Co-Creation

21. Ana Alačovska
Genre and Autonomy in Cultural
Production
The case of travel guidebook
production

22. Marius Gudmand-Høyer
Stemningssindssygdommenes historie
i det 19. århundrede
Omtydningen af melankolien og
manien som bipolære stemningslidelser
i dansk sammenhæng under hensyn til
dannelsen af det moderne følelseslivs
relative autonomi.
En problematiserings- og erfarings-
analytisk undersøgelse

23. Lichen Alex Yu
Fabricating an S&OP Process
Circulating References and Matters
of Concern

24. Esben Alfort
The Expression of a Need
Understanding search

25. Trine Pallesen
Assembling Markets for Wind Power
An Inquiry into the Making of
Market Devices

26. Anders Koed Madsen
Web-Visions
Repurposing digital traces to organize
social attention

27. Lærke Højgaard Christiansen
BREWING ORGANIZATIONAL
RESPONSES TO INSTITUTIONAL LOGICS

28. Tommy Kjær Lassen
EGENTLIG SELVLEDELSE
En ledelsesfilosofisk afhandling om
selvledelsens paradoksale dynamik og
eksistentielle engagement

29. Morten Rossing
Local Adaption and Meaning Creation in Performance Appraisal

30. Søren Obed Madsen
Lederen som oversætter
Et oversættelsesteoretisk perspektiv på strategisk arbejde

31. Thomas Høgenhaven
Open Government Communities
Does Design Affect Participation?

32. Kirstine Zinck Pedersen
Failsafe Organizing?
A Pragmatic Stance on Patient Safety

33. Anne Petersen
Hverdagslogikker i psykiatrisk arbejde
En institutionsetnografisk undersøgelse af hverdagen i psykiatriske organisationer

34. Didde Maria Humle
Fortællinger om arbejde

35. Mark Holst-Mikkelsen
Strategieksekvering i praksis
– barrierer og muligheder!

36. Malek Maalouf
Sustaining lean
Strategies for dealing with organizational paradoxes

37. Nicolaj Tofte Brenneche
Systemic Innovation In The Making
The Social Productivity of Cartographic Crisis and Transitions in the Case of SEEIT

38. Morten Gylling
The Structure of Discourse
A Corpus-Based Cross-Linguistic Study

39. Binzhang YANG
Urban Green Spaces for Quality Life
- Case Study: the landscape architecture for people in Copenhagen

40. Michael Friis Pedersen
Finance and Organization:
The Implications for Whole Farm Risk Management

41. Even Fallan
Issues on supply and demand for environmental accounting information

42. Ather Nawaz
Website user experience
A cross-cultural study of the relation between users´ cognitive style, context of use, and information architecture of local websites

43. Karin Beukel
The Determinants for Creating Valuable Inventions

44. Arjan Markus
External Knowledge Sourcing and Firm Innovation
Essays on the Micro-Foundations of Firms' Search for Innovation

2014

1. Solon Moreira
Four Essays on Technology Licensing and Firm Innovation

2. Karin Strzeletz Ivertsen
Partnership Drift in Innovation Processes
A study of the Think City electric car development

3. Kathrine Hoffmann Pii
Responsibility Flows in Patient-centred Prevention

4. Jane Bjørn Vedel
Managing Strategic Research
An empirical analysis of science-industry collaboration in a pharmaceutical company

5. Martin Gylling
Processuel strategi i organisationer
Monografi om dobbeltheden i tænkning af strategi, dels som vidensfelt i organisationsteori, dels som kunstnerisk tilgang til at skabe i erhvervsmæssig innovation

6. Linne Marie Lauesen
Corporate Social Responsibility in the Water Sector: How Material Practices and their Symbolic and Physical Meanings Form a Colonising Logic

7. Maggie Qiuzhu Mei
LEARNING TO INNOVATE: The role of ambidexterity, standard, and decision process

8. Inger Høedt-Rasmussen
Developing Identity for Lawyers Towards Sustainable Lawyering

9. Sebastian Fux
Essays on Return Predictability and Term Structure Modelling

10. Thorbjørn N. M. Lund-Poulsen
Essays on Value Based Management

11. Oana Brindusa Albu
Transparency in Organizing: A Performative Approach

12. Lena Olaison
Entrepreneurship at the limits

13. Hanne Sørum
DRESSED FOR WEB SUCCESS? An Empirical Study of Website Quality in the Public Sector

14. Lasse Folke Henriksen
Knowing networks How experts shape transnational governance

15. Maria Halbinger
Entrepreneurial Individuals Empirical Investigations into Entrepreneurial Activities of Hackers and Makers

16. Robert Spliid
Kapitalfondenes metoder og kompetencer

17. Christiane Stelling
Public-private partnerships & the need, development and management of trusting A processual and embedded exploration

18. Marta Gasparin
Management of design as a translation process

19. Kåre Moberg
Assessing the Impact of Entrepreneurship Education From ABC to PhD

20. Alexander Cole
Distant neighbors Collective learning beyond the cluster

21. Martin Møller Boje Rasmussen
Is Competitiveness a Question of Being Alike? How the United Kingdom, Germany and Denmark Came to Compete through their Knowledge Regimes from 1993 to 2007

22. Anders Ravn Sørensen
Studies in central bank legitimacy, currency and national identity Four cases from Danish monetary history

23. Nina Bellak
Can Language be Managed in International Business? Insights into Language Choice from a Case Study of Danish and Austrian Multinational Corporations (MNCs)

24. Rikke Kristine Nielsen
Global Mindset as Managerial Meta-competence and Organizational Capability: Boundary-crossing Leadership Cooperation in the MNC The Case of 'Group Mindset' in Solar A/S.

25. Rasmus Koss Hartmann
User Innovation inside government Towards a critically performative foundation for inquiry

26. Kristian Gylling Olesen
Flertydig og emergerende ledelse i folkeskolen
Et aktør-netværksteoretisk ledelses-studie af politiske evalueringsreformers betydning for ledelse i den danske folkeskole

27. Troels Riis Larsen
Kampen om Danmarks omdømme 1945-2010
Omdømmearbejde og omdømmepolitik

28. Klaus Majgaard
Jagten på autenticitet i offentlig styring

29. Ming Hua Li
Institutional Transition and Organizational Diversity:
Differentiated internationalization strategies of emerging market state-owned enterprises

30. Sofie Blinkenberg Federspiel
IT, organisation og digitalisering: Institutionelt arbejde i den kommunale digitaliseringsproces

31. Elvi Weinreich
Hvilke offentlige ledere er der brug for når velfærdstænkningen flytter sig – er Diplomuddannelsens lederprofil svaret?

32. Ellen Mølgaard Korsager
Self-conception and image of context in the growth of the firm – A Penrosian History of Fiberline Composites

33. Else Skjold
The Daily Selection

34. Marie Louise Conradsen
The Cancer Centre That Never Was
The Organisation of Danish Cancer Research 1949-1992

35. Virgilio Failla
Three Essays on the Dynamics of Entrepreneurs in the Labor Market

36. Nicky Nedergaard
Brand-Based Innovation
Relational Perspectives on Brand Logics and Design Innovation Strategies and Implementation

37. Mads Gjedsted Nielsen
Essays in Real Estate Finance

38. Kristin Martina Brandl
Process Perspectives on Service Offshoring

39. Mia Rosa Koss Hartmann
In the gray zone
With police in making space for creativity

40. Karen Ingerslev
Healthcare Innovation under The Microscope
Framing Boundaries of Wicked Problems

41. Tim Neerup Themsen
Risk Management in large Danish public capital investment programmes

2015

1. Jakob Ion Wille
Film som design
Design af levende billeder i film og tv-serier

2. Christiane Mossin
Interzones of Law and Metaphysics
Hierarchies, Logics and Foundations of Social Order seen through the Prism of EU Social Rights

3. Thomas Tøth
TRUSTWORTHINESS: ENABLING GLOBAL COLLABORATION
An Ethnographic Study of Trust, Distance, Control, Culture and Boundary Spanning within Offshore Outsourcing of IT Services

4. Steven Højlund
Evaluation Use in Evaluation Systems – The Case of the European Commission

5. Julia Kirch Kirkegaard
AMBIGUOUS WINDS OF CHANGE – OR FIGHTING AGAINST WINDMILLS IN CHINESE WIND POWER
A CONSTRUCTIVIST INQUIRY INTO CHINA'S PRAGMATICS OF GREEN MARKETISATION MAPPING CONTROVERSIES OVER A POTENTIAL TURN TO QUALITY IN CHINESE WIND POWER

6. Michelle Carol Antero
A Multi-case Analysis of the Development of Enterprise Resource Planning Systems (ERP) Business Practices

Morten Friis-Olivarius
The Associative Nature of Creativity

7. Mathew Abraham
New Cooperativism:
A study of emerging producer organisations in India

8. Stine Hedegaard
Sustainability-Focused Identity: Identity work performed to manage, negotiate and resolve barriers and tensions that arise in the process of constructing or ganizational identity in a sustainability context

9. Cecilie Glerup
Organizing Science in Society – the conduct and justification of resposible research

10. Allan Salling Pedersen
Implementering af ITIL® IT-governance - når best practice konflikter med kulturen Løsning af implementerings- problemer gennem anvendelse af kendte CSF i et aktionsforskningsforløb.

11. Nihat Misir
A Real Options Approach to Determining Power Prices

12. Mamdouh Medhat
MEASURING AND PRICING THE RISK OF CORPORATE FAILURES

13. Rina Hansen
Toward a Digital Strategy for Omnichannel Retailing

14. Eva Pallesen
In the rhythm of welfare creation
A relational processual investigation moving beyond the conceptual horizon of welfare management

15. Gouya Harirchi
In Search of Opportunities: Three Essays on Global Linkages for Innovation

16. Lotte Holck
Embedded Diversity: A critical ethnographic study of the structural tensions of organizing diversity

17. Jose Daniel Balarezo
Learning through Scenario Planning

18. Louise Pram Nielsen
Knowledge dissemination based on terminological ontologies. Using eye tracking to further user interface design.

19. Sofie Dam
PUBLIC-PRIVATE PARTNERSHIPS FOR INNOVATION AND SUSTAINABILITY TRANSFORMATION
An embedded, comparative case study of municipal waste management in England and Denmark

20. Ulrik Hartmyer Christiansen
Follwoing the Content of Reported Risk Across the Organization

21. Guro Refsum Sanden
Language strategies in multinational corporations. A cross-sector study of financial service companies and manufacturing companies.

22. Linn Gevoll
Designing performance management for operational level
- A closer look on the role of design choices in framing coordination and motivation

23. Frederik Larsen
Objects and Social Actions
– on Second-hand Valuation Practices

24. Thorhildur Hansdottir Jetzek
The Sustainable Value of Open Government Data
Uncovering the Generative Mechanisms of Open Data through a Mixed Methods Approach

25. Gustav Toppenberg
Innovation-based M&A
– Technological-Integration Challenges – The Case of Digital-Technology Companies

26. Mie Plotnikof
Challenges of Collaborative Governance
An Organizational Discourse Study of Public Managers' Struggles with Collaboration across the Daycare Area

27. Christian Garmann Johnsen
Who Are the Post-Bureaucrats?
A Philosophical Examination of the Creative Manager, the Authentic Leader and the Entrepreneur

28. Jacob Brogaard-Kay
Constituting Performance Management
A field study of a pharmaceutical company

29. Rasmus Ploug Jenle
Engineering Markets for Control: Integrating Wind Power into the Danish Electricity System

30. Morten Lindholst
Complex Business Negotiation: Understanding Preparation and Planning

31. Morten Grynings
TRUST AND TRANSPARENCY FROM AN ALIGNMENT PERSPECTIVE

32. Peter Andreas Norn
Byregimer og styringsevne: Politisk lederskab af store byudviklingsprojekter

33. Milan Miric
Essays on Competition, Innovation and Firm Strategy in Digital Markets

34. Sanne K. Hjordrup
The Value of Talent Management
Rethinking practice, problems and possibilities

35. Johanna Sax
Strategic Risk Management
– Analyzing Antecedents and Contingencies for Value Creation

36. Pernille Rydén
Strategic Cognition of Social Media

37. Mimmi Sjöklint
The Measurable Me
- The Influence of Self-tracking on the User Experience

38. Juan Ignacio Staricco
Towards a Fair Global Economic Regime? A critical assessment of Fair Trade through the examination of the Argentinean wine industry

39. Marie Henriette Madsen
Emerging and temporary connections in Quality work

40. Yangfeng CAO
Toward a Process Framework of Business Model Innovation in the Global Context
Entrepreneurship-Enabled Dynamic Capability of Medium-Sized Multinational Enterprises

41. Carsten Scheibye
Enactment of the Organizational Cost Structure in Value Chain Configuration
A Contribution to Strategic Cost Management

2016

1. Signe Sofie Dyrby
Enterprise Social Media at Work

2. Dorte Boesby Dahl
The making of the public parking attendant
Dirt, aesthetics and inclusion in public service work

3. Verena Girschik
Realizing Corporate Responsibility Positioning and Framing in Nascent Institutional Change

4. Anders Ørding Olsen
IN SEARCH OF SOLUTIONS
Inertia, Knowledge Sources and Diversity in Collaborative Problem-solving

5. Pernille Steen Pedersen
Udkast til et nyt copingbegreb
En kvalifikation af ledelsesmuligheder for at forebygge sygefravær ved psykiske problemer.

6. Kerli Kant Hvass
Weaving a Path from Waste to Value: Exploring fashion industry business models and the circular economy

7. Kasper Lindskow
Exploring Digital News Publishing Business Models – a production network approach

8. Mikkel Mouritz Marfelt
The chameleon workforce:
Assembling and negotiating the content of a workforce

9. Marianne Bertelsen
Aesthetic encounters
Rethinking autonomy, space & time in today's world of art

10. Louise Hauberg Wilhelmsen
EU PERSPECTIVES ON INTERNATIONAL COMMERCIAL ARBITRATION

11. Abid Hussain
On the Design, Development and Use of the Social Data Analytics Tool (SODATO): Design Propositions, Patterns, and Principles for Big Social Data Analytics

12. Mark Bruun
Essays on Earnings Predictability

13. Tor Bøe-Lillegraven
BUSINESS PARADOXES, BLACK BOXES, AND BIG DATA: BEYOND ORGANIZATIONAL AMBIDEXTERITY

14. Hadis Khonsary-Atighi
ECONOMIC DETERMINANTS OF DOMESTIC INVESTMENT IN AN OIL-BASED ECONOMY: THE CASE OF IRAN (1965-2010)

15. Maj Lervad Grasten
Rule of Law or Rule by Lawyers?
On the Politics of Translation in Global Governance

16. Lene Granzau Juel-Jacobsen
SUPERMARKEDETS MODUS OPERANDI – en hverdagssociologisk undersøgelse af forholdet mellem rum og handlen og understøtte relationsopbygning?

17. Christine Thalsgård Henriques
In search of entrepreneurial learning – Towards a relational perspective on incubating practices?

18. Patrick Bennett
Essays in Education, Crime, and Job Displacement

19. Søren Korsgaard
Payments and Central Bank Policy

20. Marie Kruse Skibsted
Empirical Essays in Economics of Education and Labor

21. Elizabeth Benedict Christensen
The Constantly Contingent Sense of Belonging of the 1.5 Generation Undocumented Youth
An Everyday Perspective

22. Lasse J. Jessen
Essays on Discounting Behavior and Gambling Behavior

23. Kalle Johannes Rose
Når stifterviljen dør…
Et retsøkonomisk bidrag til 200 års juridisk konflikt om ejendomsretten

24. Andreas Søeborg Kirkedal
Danish Stød and Automatic Speech Recognition

25. Ida Lunde Jørgensen
Institutions and Legitimations in Finance for the Arts

26. Olga Rykov Ibsen
An empirical cross-linguistic study of directives: A semiotic approach to the sentence forms chosen by British, Danish and Russian speakers in native and ELF contexts

27. Desi Volker
Understanding Interest Rate Volatility

28. Angeli Elizabeth Weller
Practice at the Boundaries of Business Ethics & Corporate Social Responsibility

29. Ida Danneskiold-Samsøe
Levende læring i kunstneriske organisationer
En undersøgelse af læringsprocesser mellem projekt og organisation på Aarhus Teater

30. Leif Christensen
Quality of information – The role of internal controls and materiality

31. Olga Zarzecka
Tie Content in Professional Networks

32. Henrik Mahncke
De store gaver
- Filantropiens gensidighedsrelationer i teori og praksis

33. Carsten Lund Pedersen
Using the Collective Wisdom of Frontline Employees in Strategic Issue Management

34. Yun Liu
Essays on Market Design

35. Denitsa Hazarbassanova Blagoeva
The Internationalisation of Service Firms

36. Manya Jaura Lind
Capability development in an off-shoring context: How, why and by whom

37. Luis R. Boscán F.
Essays on the Design of Contracts and Markets for Power System Flexibility

38. Andreas Philipp Distel
Capabilities for Strategic Adaptation: Micro-Foundations, Organizational Conditions, and Performance Implications

39. Lavinia Bleoca
The Usefulness of Innovation and Intellectual Capital in Business Performance: The Financial Effects of Knowledge Management vs. Disclosure

40. Henrik Jensen
Economic Organization and Imperfect Managerial Knowledge: A Study of the Role of Managerial Meta-Knowledge in the Management of Distributed Knowledge

41. Stine Mosekjær
The Understanding of English Emotion Words by Chinese and Japanese Speakers of English as a Lingua Franca An Empirical Study

42. Hallur Tor Sigurdarson
The Ministry of Desire - Anxiety and entrepreneurship in a bureaucracy

43. Kätlin Pulk
Making Time While Being in Time
A study of the temporality of organizational processes

44. Valeria Giacomin
Contextualizing the cluster Palm oil in Southeast Asia in global perspective (1880s–1970s)

45. Jeanette Willert
Managers' use of multiple Management Control Systems: The role and interplay of management control systems and company performance

46. Mads Vestergaard Jensen
Financial Frictions: Implications for Early Option Exercise and Realized Volatility

47. Mikael Reimer Jensen
Interbank Markets and Frictions

48. Benjamin Faigen
Essays on Employee Ownership

49. Adela Michea
Enacting Business Models
An Ethnographic Study of an Emerging Business Model Innovation within the Frame of a Manufacturing Company.

50. Iben Sandal Stjerne
Transcending organization in temporary systems
Aesthetics' organizing work and employment in Creative Industries

51. Simon Krogh
Anticipating Organizational Change

52. Sarah Netter
Exploring the Sharing Economy

53. Lene Tolstrup Christensen
State-owned enterprises as institutional market actors in the marketization of public service provision:
A comparative case study of Danish and Swedish passenger rail 1990–2015

54. Kyoung(Kay) Sun Park
Three Essays on Financial Economics

2017

1. Mari Bjerck
Apparel at work. Work uniforms and women in male-dominated manual occupations.

2. Christoph H. Flöthmann
Who Manages Our Supply Chains? Backgrounds, Competencies and Contributions of Human Resources in Supply Chain Management

3. Aleksandra Anna Rzeźnik
Essays in Empirical Asset Pricing

4. Claes Bäckman
Essays on Housing Markets

5. Kirsti Reitan Andersen
Stabilizing Sustainability
in the Textile and Fashion Industry

6. Kira Hoffmann
Cost Behavior: An Empirical Analysis of Determinants and Consequences of Asymmetries

7. Tobin Hanspal
Essays in Household Finance

8. Nina Lange
Correlation in Energy Markets

9. Anjum Fayyaz
Donor Interventions and SME Networking in Industrial Clusters in Punjab Province, Pakistan

10. Magnus Paulsen Hansen
Trying the unemployed. Justification and critique, emancipation and coercion towards the 'active society'. A study of contemporary reforms in France and Denmark

11. Sameer Azizi
Corporate Social Responsibility in Afghanistan
– a critical case study of the mobile telecommunications industry

12. Malene Myhre
The internationalization of small and medium-sized enterprises: A qualitative study

13. Thomas Presskorn-Thygesen
The Significance of Normativity – Studies in Post-Kantian Philosophy and Social Theory

14. Federico Clementi
Essays on multinational production and international trade

15. Lara Anne Hale
Experimental Standards in Sustainability Transitions: Insights from the Building Sector

16. Richard Pucci
Accounting for Financial Instruments in an Uncertain World Controversies in IFRS in the Aftermath of the 2008 Financial Crisis

17. Sarah Maria Denta
Kommunale offentlige private partnerskaber Regulering I skyggen af Farumsagen

18. Christian Östlund
Design for e-training

19. Amalie Martinus Hauge
Organizing Valuations – a pragmatic inquiry

20. Tim Holst Celik
Tension-filled Governance? Exploring the Emergence, Consolidation and Reconfiguration of Legitimatory and Fiscal State-crafting

21. Christian Bason
Leading Public Design: How managers engage with design to transform public governance

22. Davide Tomio
Essays on Arbitrage and Market Liquidity

23. Simone Stæhr
Financial Analysts' Forecasts Behavioral Aspects and the Impact of Personal Characteristics

24. Mikkel Godt Gregersen
Management Control, Intrinsic Motivation and Creativity – How Can They Coexist

25. Kristjan Johannes Suse Jespersen
Advancing the Payments for Ecosystem Service Discourse Through Institutional Theory

26. Kristian Bondo Hansen
Crowds and Speculation: A study of crowd phenomena in the U.S. financial markets 1890 to 1940

27. Lars Balslev
Actors and practices – An institutional study on management accounting change in Air Greenland

28. Sven Klingler
Essays on Asset Pricing with Financial Frictions

29. Klement Ahrensbach Rasmussen
Business Model Innovation The Role of Organizational Design

30. Giulio Zichella
Entrepreneurial Cognition. Three essays on entrepreneurial behavior and cognition under risk and uncertainty

31. Richard Ledborg Hansen
En forkærlighed til det eksisterende – mellemlederens oplevelse af forandringsmodstand i organisatoriske forandringer

32. Vilhelm Stefan Holsting
Militært chefvirke: Kritik og retfærdiggørelse mellem politik og profession

33. Thomas Jensen
Shipping Information Pipeline: An information infrastructure to improve international containerized shipping

34. Dzmitry Bartalevich
Do economic theories inform policy? Analysis of the influence of the Chicago School on European Union competition policy

35. Kristian Roed Nielsen
Crowdfunding for Sustainability: A study on the potential of reward-based crowdfunding in supporting sustainable entrepreneurship

36. Emil Husted
There is always an alternative: A study of control and commitment in political organization

37. Anders Ludvig Sevelsted
Interpreting Bonds and Boundaries of Obligation. A genealogy of the emergence and development of Protestant voluntary social work in Denmark as shown through the cases of the Copenhagen Home Mission and the Blue Cross (1850 – 1950)

38. Niklas Kohl
Essays on Stock Issuance

39. Maya Christiane Flensborg Jensen
BOUNDARIES OF PROFESSIONALIZATION AT WORK An ethnography-inspired study of care workers' dilemmas at the margin

40. Andreas Kamstrup
Crowdsourcing and the Architectural Competition as Organisational Technologies

41. Louise Lyngfeldt Gorm Hansen
Triggering Earthquakes in Science, Politics and Chinese Hydropower - A Controversy Study

2018

1. Vishv Priya Kohli
Combatting Falsifi cation and Counterfeiting of Medicinal Products in the E uropean Union – A Legal Analysis

2. Helle Haurum
Customer Engagement Behavior in the context of Continuous Service Relationships

3. Nis Grünberg
The Party -state order: Essays on China's political organization and political economic institutions

4. Jesper Christensen
A Behavioral Theory of Human Capital Integration

5. Poula Marie Helth
Learning in practice

6. Rasmus Vendler Toft-Kehler
Entrepreneurship as a career? An investigation of the relationship between entrepreneurial experience and entrepreneurial outcome

7. Szymon Furtak
Sensing the Future: Designing sensor-based predictive information systems for forecasting spare part demand for diesel engines

8. Mette Brehm Johansen
Organizing patient involvement. An ethnographic study

9. Iwona Sulinska
Complexities of Social Capital in Boards of Directors

10. Cecilie Fanøe Petersen
Award of public contracts as a means to conferring State aid: A legal analysis of the interface between public procurement law and State aid law

11. Ahmad Ahmad Barirani
Three Experimental Studies on Entrepreneurship

12. Carsten Allerslev Olsen
Financial Reporting Enforcement: Impact and Consequences

13. Irene Christensen
New product fumbles – Organizing for the Ramp-up process

14. Jacob Taarup-Esbensen
Managing communities – Mining MNEs' community risk management practices

15. Lester Allan Lasrado
Set-Theoretic approach to maturity models

16. Mia B. Münster
Intention vs. Perception of Designed Atmospheres in Fashion Stores

17. Anne Sluhan
Non-Financial Dimensions of Family Firm Ownership: How Socioemotional Wealth and Familiness Influence Internationalization

18. Henrik Yde Andersen
Essays on Debt and Pensions

19. Fabian Heinrich Müller
Valuation Reversed – When Valuators are Valuated. An Analysis of the Perception of and Reaction to Reviewers in Fine-Dining

20. Martin Jarmatz
Organizing for Pricing

21. Niels Joachim Christfort Gormsen
Essays on Empirical Asset Pricing

22. Diego Zunino
Socio-Cognitive Perspectives in Business Venturing

23. Benjamin Asmussen
Networks and Faces between Copenhagen and Canton, 1730-1840

24. Dalia Bagdziunaite
Brains at Brand Touchpoints A Consumer Neuroscience Study of Information Processing of Brand Advertisements and the Store Environment in Compulsive Buying

25. Erol Kazan
Towards a Disruptive Digital Platform Model

26. Andreas Bang Nielsen
Essays on Foreign Exchange and Credit Risk

27. Anne Krebs
Accountable, Operable Knowledge Toward Value Representations of Individual Knowledge in Accounting

28. Matilde Fogh Kirkegaard
A firm- and demand-side perspective on behavioral strategy for value creation: Insights from the hearing aid industry

29. Agnieszka Nowinska
SHIPS AND RELATION-SHIPS Tie formation in the sector of shipping intermediaries in shipping

30. Stine Evald Bentsen
The Comprehension of English Texts by Native Speakers of English and Japanese, Chinese and Russian Speakers of English as a Lingua Franca. An Empirical Study.

31. Stine Louise Daetz
Essays on Financial Frictions in Lending Markets

32. *Christian Skov Jensen*
Essays on Asset Pricing

33. Anders Kryger
Aligning future employee action and corporate strategy in a resource-scarce environment

34. Maitane Elorriaga-Rubio
The behavioral foundations of strategic decision-making: A contextual perspective

35. Roddy Walker
Leadership Development as Organisational Rehabilitation: Shaping Middle-Managers as Double Agents

36. Jinsun Bae
Producing Garments for Global Markets Corporate social responsibility (CSR) in Myanmar's export garment industry 2011–2015

37. Queralt Prat-i-Pubill
Axiological knowledge in a knowledge driven world. Considerations for organizations.

38. Pia Mølgaard
Essays on Corporate Loans and Credit Risk

39. Marzia Aricò
Service Design as a Transformative Force: Introduction and Adoption in an Organizational Context

40. Christian Dyrlund Wåhlin-Jacobsen
Constructing change initiatives in workplace voice activities Studies from a social interaction perspective

41. Peter Kalum Schou
Institutional Logics in Entrepreneurial Ventures: How Competing Logics arise and shape organizational processes and outcomes during scale-up

42. Per Henriksen
Enterprise Risk Management Rationaler og paradokser i en moderne ledelsesteknologi

43. Maximilian Schellmann
The Politics of Organizing Refugee Camps

44. Jacob Halvas Bjerre
Excluding the Jews: The Aryanization of Danish-German Trade and German Anti-Jewish Policy in Denmark 1937-1943

45. Ida Schrøder
Hybridising accounting and caring: A symmetrical study of how costs and needs are connected in Danish child protection work

46. Katrine Kunst
Electronic Word of Behavior: Transforming digital traces of consumer behaviors into communicative content in product design

47. Viktor Avlonitis
Essays on the role of modularity in management: Towards a unified perspective of modular and integral design

48. Anne Sofie Fischer
Negotiating Spaces of Everyday Politics: -An ethnographic study of organizing for social transformation for women in urban poverty, Delhi, India

2019

1. Shihan Du
ESSAYS IN EMPIRICAL STUDIES BASED ON ADMINISTRATIVE LABOUR MARKET DATA

2. Mart Laatsit
Policy learning in innovation policy: A comparative analysis of European Union member states

3. Peter J. Wynne
Proactively Building Capabilities for the Post-Acquisition Integration of Information Systems

4. Kalina S. Staykova
Generative Mechanisms for Digital Platform Ecosystem Evolution

5. Ieva Linkeviciute
Essays on the Demand-Side Management in Electricity Markets

6. Jonatan Echebarria Fernández
Jurisdiction and Arbitration Agreements in Contracts for the Carriage of Goods by Sea – Limitations on Party Autonomy

7. Louise Thorn Bøttkjær
Votes for sale. Essays on clientelism in new democracies.

8. Ditte Vilstrup Holm
The Poetics of Participation: the organizing of participation in contemporary art

9. Philip Rosenbaum
Essays in Labor Markets – Gender, Fertility and Education

10. Mia Olsen
Mobile Betalinger - Succesfaktorer og Adfærdsmæssige Konsekvenser

11. Adrián Luis Mérida Gutiérrez
Entrepreneurial Careers: Determinants, Trajectories, and Outcomes

12. Frederik Regli
Essays on Crude Oil Tanker Markets

13. Cancan Wang
Becoming Adaptive through Social Media: Transforming Governance and Organizational Form in Collaborative E-government

14. Lena Lindbjerg Sperling
Economic and Cultural Development: Empirical Studies of Micro-level Data

15. Xia Zhang
Obligation, face and facework: An empirical study of the communicative act of cancellation of an obligation by Chinese, Danish and British business professionals in both L1 and ELF contexts

16. Stefan Kirkegaard Sløk-Madsen
Entrepreneurial Judgment and Commercialization

17. Erin Leitheiser
The Comparative Dynamics of Private Governance
The case of the Bangladesh Ready-Made Garment Industry

18. Lone Christensen
STRATEGIIMPLEMENTERING: STYRINGSBESTRÆBELSER, IDENTITET OG AFFEKT

19. Thomas Kjær Poulsen
Essays on Asset Pricing with Financial Frictions

20. Maria Lundberg
Trust and self-trust in leadership identity constructions: A qualitative exploration of narrative ecology in the discursive aftermath of heroic discourse

21. Tina Joanes
Sufficiency for sustainability
Determinants and strategies for reducing clothing consumption

22. Benjamin Johannes Flesch
Social Set Visualizer (SoSeVi): Design, Development and Evaluation of a Visual Analytics Tool for Computational Set Analysis of Big Social Data

23. Henriette Sophia Groskopff Tvede Schleimann
Creating innovation through collaboration – Partnering in the maritime sector

24. Kristian Steensen Nielsen
The Role of Self-Regulation in Environmental Behavior Change

25. Lydia L. Jørgensen
Moving Organizational Atmospheres

26. Theodor Lucian Vladasel
Embracing Heterogeneity: Essays in Entrepreneurship and Human Capital

27. Seidi Suurmets
Contextual Effects in Consumer Research: An Investigation of Consumer Information Processing and Behavior via the Applicati on of Eye-tracking Methodology

28. Marie Sundby Palle Nickelsen
Reformer mellem integritet og innovation: Reform af reformens form i den danske centraladministration fra 1920 til 2019

29. Vibeke Kristine Scheller
The temporal organizing of same-day discharge: A tempography of a Cardiac Day Unit

30. Qian Sun
Adopting Artificial Intelligence in Healthcare in the Digital Age: Perceived Challenges, Frame Incongruence, and Social Power

31. Dorthe Thorning Mejlhede
Artful change agency and organizing for innovation – the case of a Nordic fintech cooperative

32. Benjamin Christoffersen
Corporate Default Models:
Empirical Evidence and Methodical Contributions

33. Filipe Antonio Bonito Vieira
Essays on Pensions and Fiscal Sustainability

34. Morten Nicklas Bigler Jensen
Earnings Management in Private Firms: An Empirical Analysis of Determinants and Consequences of Earnings Management in Private Firms

2020

1. Christian Hendriksen
Inside the Blue Box: Explaining industry influence in the International Maritime Organization

2. Vasileios Kosmas
Environmental and social issues in global supply chains:
Emission reduction in the maritime transport industry and maritime search and rescue operational response to migration

3. Thorben Peter Simonsen
The spatial organization of psychiatric practice: A situated inquiry into 'healing architecture'

4. Signe Bruskin
The infinite storm: An ethnographic study of organizational change in a bank

5. Rasmus Corlin Christensen
Politics and Professionals: Transnational Struggles to Change International Taxation

6. Robert Lorenz Törmer
The Architectural Enablement of a Digital Platform Strategy

7. Anna Kirkebæk Johansson Gosovic
Ethics as Practice: An ethnographic study of business ethics in a multinational biopharmaceutical company

8. Frank Meier
Making up leaders in leadership development

9. Kai Basner
Servitization at work: On proliferation and containment

10. Anestis Keremis
Anti-corruption in action: How is anti-corruption practiced in multinational companies?

11. Marie Larsen Ryberg
Governing Interdisciolinarity: Stakes and translations of interdisciplinarity in Danish high school education.

12. Jannick Friis Christensen
Queering organisation(s): Norm-critical orientations to organising and researching diversity

13. Thorsteinn Sigurdur Sveinsson
Essays on Macroeconomic Implications of Demographic Change

14. Catherine Casler
Reconstruction in strategy and organization: For a pragmatic stance

15. Luisa Murphy
Revisiting the standard organization of multi-stakeholder initiatives (MSIs): The case of a meta-MSI in Southeast Asia

16. Friedrich Bergmann
Essays on International Trade

17. Nicholas Haagensen
European Legal Networks in Crisis: The Legal Construction of Economic Policy

18. Charlotte Biil
Samskabelse med en sommerfugle-model: Hybrid ret i forbindelse med et partnerskabsprojekt mellem 100 selvejende daginstitutioner, deres paraplyorganisation, tre kommuner og CBS

19. Andreas Dimmelmeier
The Role of Economic Ideas in Sustainable Finance: From Paradigms to Policy

20. Maibrith Kempka Jensen
Ledelse og autoritet i interaktion - En interaktionsbaseret undersøgelse af autoritet i ledelse i praksis

21. Thomas Burø
LAND OF LIGHT: Assembling the Ecology of Culture in Odsherred 2000-2018

22. Prins Marcus Valiant Lantz
Timely Emotion: The Rhetorical Framing of Strategic Decision Making

23. Thorbjørn Vittenhof Fejerskov
Fra værdi til invitationer - offentlig værdiskabelse gennem affekt, potentialitet og begivenhed

24. Lea Acre Foverskov
Demographic Change and Employment: Path dependencies and institutional logics in the European Commission

25. Anirudh Agrawal
A Doctoral Dissertation

26. Julie Marx
Households in the housing market

27. Hadar Gafni
Alternative Digital Methods of Providing Entrepreneurial Finance

28. Mathilde Hjerrild Carlsen
Ledelse af engagementer: En undersøgelse af samarbejde mellem folkeskoler og virksomheder i Danmark

29. Suen Wang
Essays on the Gendered Origins and Implications of Social Policies in the Developing World

30. Stine Hald Larsen
The Story of the Relative: A Systems-Theoretical Analysis of the Role of the Relative in Danish Eldercare Policy from 1930 to 2020

31. Christian Casper Hofma
Immersive technologies and organizational routines: When head-mounted displays meet organizational routines

32. Jonathan Feddersen
The temporal emergence of social relations: An event-based perspective of organising

33. Nageswaran Vaidyanathan
ENRICHING RETAIL CUSTOMER EXPERIENCE USING AUGMENTED REALITY

2021

1. Vanya Rusinova
The Determinants of Firms' Engagement in Corporate Social Responsibility: Evidence from Natural Experiments

2. Lívia Lopes Barakat
Knowledge management mechanisms at MNCs: The enhancing effect of absorptive capacity and its effects on performance and innovation

3. Søren Bundgaard Brøgger
Essays on Modern Derivatives Markets

4. Martin Friis Nielsen
Consuming Memory: Towards a conceptualization of social media platforms as organizational technologies of consumption

05. Fei Liu
Emergent Technology Use in Consumer Decision Journeys: A Process-as-Propensity Approach

06. Jakob Rømer Barfod
Ledelse i militære højrisikoteams

07. Elham Shafiei Gol
Creative Crowdwork Arrangements

08. Árni Jóhan Petersen
Collective Imaginary as (Residual) Fantasy: A Case Study of the Faroese Oil Bonanza

09. Søren Bering
"Manufacturing, Forward Integration and Governance Strategy"

10. Lars Oehler
Technological Change and the Decomposition of Innovation: Choices and Consequences for Latecomer Firm Upgrading: The Case of China's Wind Energy Sector

11. Lise Dahl Arvedsen
Leadership in interaction in a virtual context: A study of the role of leadership processes in a complex context, and how such processes are accomplished in practice

12. Jacob Emil Jeppesen
Essays on Knowledge networks, scientific impact and new knowledge adoption

13. Kasper Ingeman Beck
Essays on Chinese State-Owned Enterprises: Reform, Corporate Governance and Subnational Diversity

14. Sönnich Dahl Sönnichsen
Exploring the interface between public demand and private supply for implementation of circular economy principles

15. Benjamin Knox
Essays on Financial Markets and Monetary Policy

16. Anita Eskesen
Essays on Utility Regulation: Evaluating Negotiation-Based Approaches inthe Context of Danish Utility Regulation

17. Agnes Guenther
Essays on Firm Strategy and Human Capital

18. Sophie Marie Cappelen
Walking on Eggshells: The balancing act of temporal work in a setting of culinary change

19. Manar Saleh Alnamlah
About Gender Gaps in Entrepreneurial Finance

20. Kirsten Tangaa Nielsen
Essays on the Value of CEOs and Directors

21. Renée Ridgway
Re:search - the Personalised Subject vs. the Anonymous User

22. Codrina Ana Maria Lauth
IMPACT Industrial Hackathons: Findings from a longitudinal case study on short-term vs long-term IMPACT implementations from industrial hackathons within Grundfos

23. Wolf-Hendrik Uhlbach
Scientist Mobility: Essays on knowledge production and innovation

24. Tomaz Sedej
Blockchain technology and inter-organizational relationships

25. Lasse Bundgaard
Public Private Innovation Partnerships: Creating Public Value & Scaling Up Sustainable City Solutions

26. Dimitra Makri Andersen
Walking through Temporal Walls: Rethinking NGO Organizing for Sustainability through a Temporal Lens on NGO-Business Partnerships

27. Louise Fjord Kjærsgaard
Allocation of the Right to Tax Income from Digital Products and Services: A legal analysis of international tax treaty law

28. Sara Dahlman
Marginal alternativity: Organizing for sustainable investing

29. Henrik Gundelach
Performance determinants: An Investigation of the Relationship between Resources, Experience and Performance in Challenging Business Environments

30. Tom Wraight
Confronting the Developmental State: American Trade Policy in the Neoliberal Era

31. Mathias Fjællegaard Jensen
Essays on Gender and Skills in the Labour Market

32. Daniel Lundgaard
Using Social Media to Discuss Global Challenges: Case Studies of the Climate Change Debate on Twitter

33. Jonas Sveistrup Søgaard
Designs for Accounting Information Systems using Distributed Ledger Technology

34. Sarosh Asad
CEO narcissism and board composition: Implications for firm strategy and performance

35. Johann Ole Willers
Experts and Markets in Cybersecurity On Definitional Power and the Organization of Cyber Risks

36. Alexander Kronies
Opportunities and Risks in Alternative Investments

37. Niels Fuglsang
The Politics of Economic Models: An inquiry into the possibilities and limits concerning the rise of macroeconomic forecasting models and what this means for policymaking

38. David Howoldt
Policy Instruments and Policy Mixes for Innovation: Analysing Their Relation to Grand Challenges, Entrepreneurship and Innovation Capability with Natural Language Processing and Latent Variable Methods

2022

01. Ditte Thøgersen
Managing Public Innovation on the Frontline

02. Rasmus Jørgensen
Essays on Empirical Asset Pricing and Private Equity

03. Nicola Giommetti
Essays on Private Equity

04. Laila Starr
When Is Health Innovation Worth It? Essays On New Approaches To Value Creation In Health

05. Maria Krysfeldt Rasmussen
Den transformative ledelsesbyrde – etnografisk studie af en religionsinspireret ledelsesfilosofi i en dansk modevirksomhed

06. Rikke Sejer Nielsen
Mortgage Decisions of Households: Consequences for Consumption and Savings

07. Myriam Noémy Marending
Essays on development challenges of low income countries: Evidence from conflict, pest and credit

08. Selorm Agbleze
A BEHAVIORAL THEORY OF FIRM FORMALIZATION

09. Rasmus Arler Bogetoft
Rettighedshavers faktisk lidte tab i immaterialretssager: Studier af dansk ret med støtte i økonomisk teori og metode

10. Franz Maximilian Buchmann
Driving the Green Transition of the Maritime Industry through Clean Technology Adoption and Environmental Policies

11. Ivan Olav Vulchanov
The role of English as an organisational language in international workplaces

12. Anne Agerbak Bilde
TRANSFORMATIONER AF SKOLELEDELSE - en systemteoretisk analyse af hvordan betingelser for skoleledelse forandres med læring som genstand i perioden 1958-2020

13. JUAN JOSE PRICE ELTON
EFFICIENCY AND PRODUCTIVITY ANALYSIS: TWO EMPIRICAL APPLICATIONS AND A METHODOLOGICAL CONTRIBUTION

14. Catarina Pessanha Gomes
The Art of Occupying: Romanticism as Political Culture in French Prefigurative politics

15. Mark Ørberg
Fondsretten og den levende vedtægt

16. Majbritt Greve
Maersk's Role in Economic Development: A Study of Shipping and Logistics Foreign Direct Investment in Global Trade

17. Sille Julie J. Abildgaard
Doing-Being Creative: Empirical Studies of Interaction in Design Work

18. Jette Sandager
Glitter, Glamour, and the Future of (More) Girls in STEM: Gendered Formations of STEM Aspirations

19. Casper Hein Winther
Inside the innovation lab - How paradoxical tensions persist in ambidextrous organizations over time

20. Nikola Kostić
Collaborative governance of inter-organizational relationships: The effects of management controls, blockchain technology, and industry standards

21. Saila Naomi Stausholm
Maximum capital, minimum tax: Enablers and facilitators of corporate tax minimization

22. Robin Porsfelt
Seeing through Signs: On Economic Imagination and Semiotic Speculation

23. Michael Herburger
Supply chain resilience – a concept for coping with cyber risks

24. Katharina Christiane Nielsen Jeschke
Balancing safety in everyday work - A case study of construction managers' dynamic safety practices

25. Jakob Ahm Sørensen
Financial Markets with Frictions and Belief Distortions

26. Jakob Laage-Thomsen
Nudging Leviathan, Protecting Demos - A Comparative Sociology of Public Administration and Expertise in the Nordics

27. Kathrine Søs Jacobsen Cesko
Collaboration between Economic Operators in the Competition for Public Contracts: A Legal and Economic Analysis of Grey Zones between EU Public Procurement Law and EU Competition Law

28. Mette Nelund
Den nye jord – Et feltstudie af et bæredygtigt virke på Farendløse Mosteri

29. Benjamin Cedric Larsen
Governing Artificial Intelligence – Lessons from the United States and China

30. Anders Brøndum Klein
Kollektiv meningsdannelse iblandt heterogene aktører i eksperimentelle samskabelsesprocesser

31. Stefano Tripodi
Essays on Development Economicis

32. Katrine Maria Lumbye
Internationalization of European Electricity Multinationals in Times of Transition

33. Xiaochun Guo
Dynamic Roles of Digital Currency – An Exploration from Interactive Processes: Difference, Time, and Perspective

34. Louise Lindbjerg
Three Essays on Firm Innovation

35. Marcela Galvis Restrepo
Feature reduction for classification with mixed data: an algorithmic approach

36. Hanna Nyborg Storm
Cultural institutions and attractiveness – How cultural institutions contribute to the development of regions and local communities

37. Anna-Bertha Heeris Christensen
Conflicts and Challenges in Practices of Commercializing Humans – An Ethnographic Study of Influencer Marketing Work

38. Casper Berg Lavmand Larsen
A Worker-Centered Inquiry into the Contingencies and Consequences of Worker Representation

39. Niels le Duc
The Resource Commitment of Multinational Enterprise R&D Activities

40. Esben Langager Olsen
Change management tools and change managers – Examining the simulacra of change

41. Anne Sophie Lassen
Gender in the Labor Market

42. Alison E. Holm
Corrective corporate responses to accusations of misconduct on societal issues

43. Chenyan Lyu
Carbon Pricing, Renewable Energy, and Clean Growth – A Market Perspective

44. Alina Grecu
UNPACKING MULTI-LEVEL OFFSHORING CONSEQUENCES: Hiring Wages, Onshore Performance, and Public Sentiment

45. Alexandra Lüth
Offshore Energy Hubs as an Emerging Concept – Sector Integration at Sea

2023

01. Cheryl Basil Sequeira
Port Business Development – Digitalisation of Port Authroity and Hybrid Governance Model

02. Mette Suder Franck
Empirical Essays on Technology Supported Learning – Studies of Danish Higher Education

03. Søren Lund Frandsen
States and Experts – Assembling Expertise for Climate Change and Pandemics

04. Guowei Dong
Innovation and Internationalization – Evidence from Chinese Manufacturing Enterprises

05. Eileen Murphy
In Service to Security – Constructing the Authority to Manage European Border Data Infrastructures

06. Bontu Lucie Guschke
THE PERSISTENCE OF SEXISM AND RACISM AT UNIVERSITIES – Exploring the imperceptibility and unspeakability of workplace harassment and discrimination in academia

07. Christoph Viebig
Learning Entrepreneurship – How capabilities shape learning from experience, reflection, and action

08. Kasper Regenburg
Financial Risks of Private Firms

09. Kathrine Møller Solgaard
Who to hire? – A situated study of employee selection as routine, practice, and process

10. Jack Kværnø-Jones
Intersections between FinTech Imaginaries and Traditional Banking – A study of disciplinary, implementary, and parasitic work in the Danish financial sector

11. Stine Quorning
Managing Climate Change Like a Central Banker – The Political Economy of Greening the Monetary Technocracy

12. Amanda Bille
No business without politics – Investigating the political nature of supply chain management

13. Theis Ingerslev Jensen
Essays on Empirical Asset Pricing

14. Ann Fugl-Meyer
The Agile Imperative – A Qualitative Study of a Translation Process in the Danish Tax Administration

15. Nicolai Søgaard Laursen
Longevity risk in reinsurance and equity markets

16. Shelter Selorm Kwesi Teyi
STRATEGIC ENTREPRENEURSHIP IN THE INFORMAL ECONOMY

17. Luisa Hedler
Time, Law and Tech – The introduction of algorithms to courts of law

18. Tróndur Møller Sandoy
Essays on the Economics of Education

TITLER I ATV PH.D.-SERIEN

1992

1. Niels Kornum
Servicesamkørsel – organisation, økonomi og planlægningsmetode

1995

2. Verner Worm
Nordiske virksomheder i Kina
Kulturspecifikke interaktionsrelationer ved nordiske virksomhedsetableringer i Kina

1999

3. Mogens Bjerre
Key Account Management of Complex Strategic Relationships
An Empirical Study of the Fast Moving Consumer Goods Industry

2000

4. Lotte Darsø
Innovation in the Making
Interaction Research with heterogeneous Groups of Knowledge Workers creating new Knowledge and new Leads

2001

5. Peter Hobolt Jensen
Managing Strategic Design Identities
The case of the Lego Developer Network

2002

6. Peter Lohmann
The Deleuzian Other of Organizational Change – Moving Perspectives of the Human

7. Anne Marie Jess Hansen
To lead from a distance: The dynamic interplay between strategy and strategizing – A case study of the strategic management process

2003

8. Lotte Henriksen
Videndeling
– om organisatoriske og ledelsesmæssige udfordringer ved videndeling i praksis

9. Niels Christian Nickelsen
Arrangements of Knowing: Coordinating Procedures Tools and Bodies in Industrial Production – a case study of the collective making of new products

2005

10. Carsten Ørts Hansen
Konstruktion af ledelsesteknologier og effektivitet

TITLER I DBA PH.D.-SERIEN

2007

1. Peter Kastrup-Misir
Endeavoring to Understand Market Orientation – and the concomitant co-mutation of the researched, the re searcher, the research itself and the truth

2009

1. Torkild Leo Thellefsen
Fundamental Signs and Significance effects
A Semeiotic outline of Fundamental Signs, Significance-effects, Knowledge Profiling and their use in Knowledge Organization and Branding

2. Daniel Ronzani
When Bits Learn to Walk Don't Make Them Trip. Technological Innovation and the Role of Regulation by Law in Information Systems Research: the Case of Radio Frequency Identification (RFID)

2010

1. Alexander Carnera
Magten over livet og livet som magt
Studier i den biopolitiske ambivalens